OM of: US-09-303-518D-653 to: Issued_Patents_AA:* out_format : pfs

Date: Jun 30, 2002 7:30 AM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09303518/runat_28062002_142713_4302/app_query.fasta_1.23501
-DB=Issued_Patents_AA -QFMT=fastan -SUFFIX=rai -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=100 -DOCALIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=45 -MODE=LOCAL -OUTFMT=pfs
-NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09303518_@CGN1_1_279 -NCPU=6 -ICPU=3 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-09-303-518D-653
Query length: 4407
Database: Issued_Patents_AA:*
Database sequences: 231628
Database length: 24425594
Search time (sec): 185.890000

score_list:

Sequence	Strd	Orig	ZScore	EScore	Len	! Documentation
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-08-296-791-3 +		1614.00	2199.41	3.2e-116	1541	!
/cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:PCT-US95-10661A-3 +		1614.00	2199.41	3.2e-116	1541	
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-08-296-791-4 +		1587.00	2161.98	3.9e-114	1545	!
/cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:PCT-US95-10661A-4 +		1587.00	2161.98	3.9e-114	1545	
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-08-296-791-5 +		1581.50	2153.24	1.1e-113	1702	!
/cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:PCT-US95-10661A-5 +		1581.50	2153.24	1.1e-113	1702	
/cgn2_6/ptodata/1/iaa/backfiles1.pep:5268270-2 +		1515.00	2062.54	1.4e-108	1507	! Pate
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-08-296-791-6 +		1513.00	2057.41	2.2e-108	1848	!
/cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:PCT-US95-10661A-6 +		1513.00	2057.41	2.2e-108	1848	
/cgn2_6/ptodata/1/iaa/5B_COMB.pep:US-08-728-470-9 +		308.50	392.74	1.6e-15	1338	!
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-08-719-641-9 +		308.50	392.74	1.6e-15	1338	!
/cgn2_6/ptodata/1/iaa/5B_COMB.pep:US-08-617-697-9 +		308.50	390.68	1.7e-15	1599	!
/cgn2_6/ptodata/1/iaa/5B_COMB.pep:US-08-617-697-10 +		301.50	380.97	6.0e-15	1600	!
/cgn2_6/ptodata/1/iaa/5B_COMB.pep:US-08-728-470-10 +		292.50	369.03	2.9e-14	1529	!
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-08-719-641-10 +		292.50	369.03	2.9e-14	1529	!
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-08-460-269C-6 +		275.50	351.33	4.7e-13	922	!
/cgn2_6/ptodata/1/iaa/5A_COMB.pep:US-08-038-682-4 +		274.50	344.50	7.0e-13	1477	!
/cgn2_6/ptodata/1/iaa/5A_COMB.pep:US-08-302-832-4 +		274.50	344.50	7.0e-13	1477	!
/cgn2_6/ptodata/1/iaa/5B_COMB.pep:US-08-530-198-4 +		274.50	344.50	7.0e-13	1477	!
/cgn2_6/ptodata/1/iaa/5B_COMB.pep:US-08-469-880-4 +		274.50	344.50	7.0e-13	1477	!
/cgn2_6/ptodata/1/iaa/5B_COMB.pep:US-08-728-470-4 +		274.50	344.50	7.0e-13	1477	!
/cgn2_6/ptodata/1/iaa/5B_COMB.pep:US-08-617-697-4 +		274.50	344.50	7.0e-13	1477	!
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-08-719-641-4 +		274.50	344.50	7.0e-13	1477	!
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-08-460-269C-4 +		274.00	349.39	6.1e-13	911	!
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-08-460-269C-2 +		268.00	341.09	1.8e-12	910	!
/cgn2_6/ptodata/1/iaa/5A_COMB.pep:US-08-038-682-2 +		265.00	330.89	3.9e-12	1536	!
/cgn2_6/ptodata/1/iaa/5A_COMB.pep:US-08-302-832-2 +		265.00	330.89	3.9e-12	1536	!
/cgn2_6/ptodata/1/iaa/5B_COMB.pep:US-08-530-198-2 +		265.00	330.89	3.9e-12	1536	!
/cgn2_6/ptodata/1/iaa/5B_COMB.pep:US-08-469-880-2 +		265.00	330.89	3.9e-12	1536	!
/cgn2_6/ptodata/1/iaa/5B_COMB.pep:US-08-728-470-2 +		265.00	330.89	3.9e-12	1536	!
/cgn2_6/ptodata/1/iaa/5B_COMB.pep:US-08-617-697-2 +		265.00	330.89	3.9e-12	1536	!
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-08-719-641-2 +		265.00	330.89	3.9e-12	1536	!
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-09-268-347-36 +		257.50	315.29	1.8e-11	2411	!
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-09-268-347-47 +		254.50	311.41	3.1e-11	2354	!
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-09-377-155-33 +		253.50	310.03	3.7e-11	2353	!
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-08-913-942-4 +		253.50	310.03	3.7e-11	2353	!
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-09-669-974-33 +		253.50	310.03	3.7e-11	2353	!
/cgn2_6/ptodata/1/iaa/5A_COMB.pep:US-08-409-995-4 +		248.50	305.50	8.1e-11	1912	!
/cgn2_6/ptodata/1/iaa/6A_COMB.pep:US-08-685-467-4 +		248.50	305.50	8.1e-11	1912	!
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-09-413-814-78 +		245.00	309.60	1.0e-10	882	!
/cgn2_6/ptodata/1/iaa/5A_COMB.pep:US-08-169-927-2 +		243.50	300.55	1.8e-10	1612	
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-09-268-347-49 +		237.00	287.37	6.8e-10	2314	
/cgn2_6/ptodata/1/iaa/6A_COMB.pep:US-08-968-685A-10 +		228.50	276.59	3.0e-09	2123	
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-09-268-347-48 +		227.00	274.93	3.8e-09	2048	
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-09-413-814-79 +		211.00	258.82	5.1e-08	1213	
/cgn2_6/ptodata/1/iaa/5A_COMB.pep:US-08-409-995-2 +		202.50	248.20	2.2e-07	1098	
/cgn2_6/ptodata/1/iaa/6A_COMB.pep:US-08-685-467-2 +		202.50	248.20	2.2e-07	1098	
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-09-377-155-32 +		202.50	248.20	2.2e-07	1098	
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-08-913-942-2 +		202.50	248.20	2.2e-07	1098	
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-09-669-974-32 +		202.50	248.20	2.2e-07	1098	
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-09-268-347-44 +		202.50	248.20	2.2e-07	1098	
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-09-199-637A-351 +		195.00	248.19	5.4e-07	447	
/cgn2_6/ptodata/1/iaa/5B_COMB.pep:US-08-387-942C-3 +		193.50	232.90	1.2e-06	1403	
/cgn2_6/ptodata/1/iaa/5B_COMB.pep:US-08-682-517-9 +		193.00	233.52	1.3e-06	1252	
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-09-268-347-28 +		191.00	232.20	1.7e-06	1104	
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-09-268-347-34 +		191.00	232.20	1.7e-06	1104	
/cgn2_6/ptodata/1/iaa/5A_COMB.pep:US-07-741-940-2 +		190.50	220.58	2.9e-06	2843	
/cgn2_6/ptodata/1/iaa/5A_COMB.pep:US-08-289-548A-2 +		190.50	220.58	2.9e-06	2843	
/cgn2_6/ptodata/1/iaa/5A_COMB.pep:US-08-452-654-2 +		190.50	220.58	2.9e-06	2843	
/cgn2_6/ptodata/1/iaa/5B_COMB.pep:US-08-370-235A-2 +		190.50	220.58	2.9e-06	2843	
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-09-268-347-32 +		188.00	228.15	2.9e-06	1094	
/cgn2_6/ptodata/1/iaa/5B_COMB.pep:US-08-682-517-15 +		187.50	226.18	3.3e-06	1222	
/cgn2_6/ptodata/1/iaa/5A_COMB.pep:US-08-200-232-2 +		187.00	224.89	3.7e-06	1287	
/cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:PCT-US95-02219-2 +		187.00	224.89	3.7e-06	12	
/cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:PCT-US95-02219A-2 +		187.00	224.89	3.7e-06	1	
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-09-268-347-30 +		186.00	226.38	3.9e-06	1004	
/cgn2_6/ptodata/1/iaa/5A_COMB.pep:US-07-741-940-7 +		185.50	213.66	7.1e-06	2842	
/cgn2_6/ptodata/1/iaa/5A_COMB.pep:US-08-289-548A-7 +		185.50	213.66	7.1e-06	2842	
/cgn2_6/ptodata/1/iaa/5A_COMB.pep:US-08-452-654-7 +		185.50	213.66	7.1e-06	2842	
/cgn2_6/ptodata/1/iaa/5A_COMB.pep:US-08-452-655B-2 +		185.50	213.65	7.1e-06	2843	
/cgn2_6/ptodata/1/iaa/5A_COMB.pep:US-08-452-655B-7 +		185.50	213.65	7.1e-06	2843	
/cgn2_6/ptodata/1/iaa/6A_COMB.pep:US-08-450-582-2 +		185.50	213.65	7.1e-06	2843	
/cgn2_6/ptodata/1/iaa/6A_COMB.pep:US-08-450-582-7 +		185.50	213.65	7.1e-06	2843	
/cgn2_6/ptodata/1/iaa/5B_COMB.pep:US-08-821-355A-7 +		185.50	213.14	7.2e-06	2973	
/cgn2_6/ptodata/1/iaa/5B_COMB.pep:US-09-003-687A-7 +		185.50	213.14	7.2e-06	2973	
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-09-136-605-7 +		185.50	213.14	7.2e-06	2973	
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-09-336-447A-15 +		178.50	217.39	1.4e-05	889	
/cgn2_6/ptodata/1/iaa/6A_COMB.pep:US-08-470-260-3 +		178.50	213.04	1.7e-05	1296	
/cgn2_6/ptodata/1/iaa/6A_COMB.pep:US-08-471-491-3 +		178.50	213.04	1.7e-05	1296	
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-08-466-662-3 +		178.50	213.04	1.7e-05	1296	
/cgn2_6/ptodata/1/iaa/5B_COMB.pep:US-08-348-353-17 +		178.00	212.78	1.8e-05	1248	
/cgn2_6/ptodata/1/iaa/5B_COMB.pep:US-08-465-965-17 +		178.00	212.78	1.8e-05	1248	
/cgn2_6/ptodata/1/iaa/6A_COMB.pep:US-08-465-966-17 +		178.00	212.78	1.8e-05	1248	
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-09-540-245A-16 +		176.50	209.53	2.5e-05	1381	
/cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:PCT-US95-11684-2 +		176.00	203.46	3.4e-05	21	
/cgn2_6/ptodata/1/iaa/5B_COMB.pep:US-08-542-003-6 +		174.50	210.20	3.1e-05	1026	
/cgn2_6/ptodata/1/iaa/5B_COMB.pep:US-08-322-760A-6 +		174.50	210.20	3.1e-05	1026	
/cgn2_6/ptodata/1/iaa/5B_COMB.pep:US-08-387-942C-4 +		173.00	208.45	3.9e-05	997	
/cgn2_6/ptodata/1/iaa/5B_COMB.pep:US-08-387-942C-5 +		172.50	209.31	4.0e-05	872	
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-09-268-347-24 +		171.50	206.31	5.1e-05	1002	
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-09-074-658-11 +		171.00	195.34	8.6e-05	2439	
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-09-336-447A-13 +		170.50	206.52	5.8e-05	873	
/cgn2_6/ptodata/1/iaa/backfiles1.pep:5352450-2 +		170.50	199.78	7.6e-05	1565	! P
/cgn2_6/ptodata/1/iaa/5A_COMB.pep:US-08-480-604A-6 +		170.00	192.74	0.0001	2710	
/cgn2_6/ptodata/1/iaa/5B_COMB.pep:US-08-405-496A-6 +		170.00	192.74	0.0001	2710	
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-08-915-136-6 +		170.00	192.74	0.0001	2710	
/cgn2_6/ptodata/1/iaa/5B_COMB.pep:US-08-616-844-40 +		169.00	198.34	9.7e-05	1481	
/cgn2_6/ptodata/1/iaa/5B_COMB.pep:US-08-599-654-40 +		169.00	198.34	9.7e-05	1481	

seq_name: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-08-296-791-3

seq_documentation_block:
; Sequence 3, Application US/08296791
; Patent No. 6245337
; GENERAL INFORMATION:
; APPLICANT: St. Geme III, Joseph W.
; APPLICANT: Falkow, Stanley
; TITLE OF INVENTION: Haemophilus Adherence and Penetration
; TITLE OF INVENTION: Protein
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
; STREET: 4 Embarcadero Center, Suite 3400

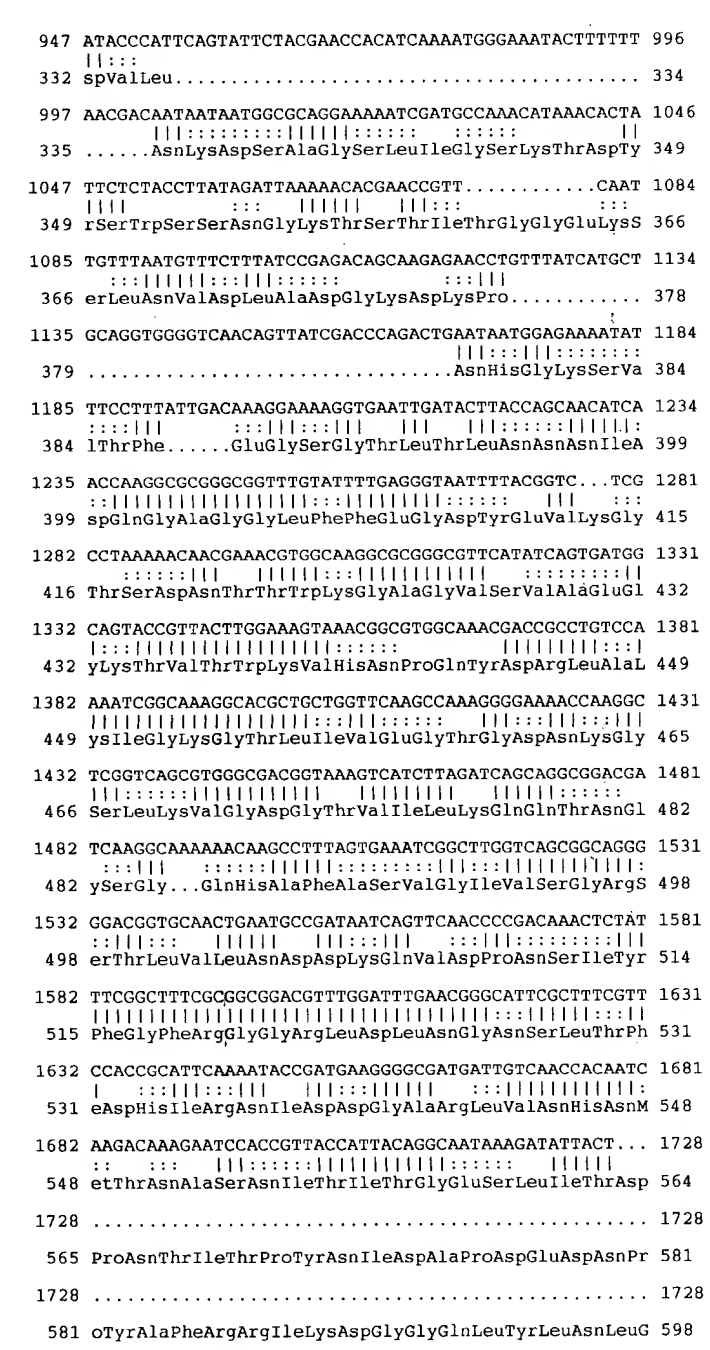
```
;       CITY:  San Francisco
;       STATE:  California
;       COUNTRY:  United States
;       ZIP:  94111-4187
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/296,791
;       FILING DATE:  25-AUG-1994
;       CLASSIFICATION:  435
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Trecartin, Richard F.
;       REGISTRATION NUMBER:  31,801
;       REFERENCE/DOCKET NUMBER:  A-59941/RFT/RMS
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (415) 781-1989
;       TELEFAX:  (415) 398-3249
;       TELEX:  910 277299
;   INFORMATION FOR SEQ ID NO:  3:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1541 amino acids
;       TYPE:  amino acid
;       TOPOLOGY:  unknown
US-08-296-791-3


alignment_scores:
            Quality: 1614.00              Length:   1716
              Ratio:  1.732                Gaps:     48
 Percent Similarity:  54.312   Percent Identity: 26.865

alignment_block:
 US-09-303-518D-653 x US-08-296-791-3   ..

 Align seg 1/1  to: US-08-296-791-3  from: 1  to: 1541

      64 CGCTTCTCGCCCGCTTACTTAGCCATATGCCTGTCGTTCGGCATTCTGCC 113
         :::|||:::     ::::::|||:::   :::::::::::::::   ||
       5 LysPheLysLeuAsnPheIleAlaLeuThrValAlaTyrAlaLeuThrPr 21

     114 CCAAGCCCGGGCGGGACACACTTATTTCGGCATCAACTACCAATACTATC 163
         |   :::::::|||:::   :::         :::::::||||||   :::|
      21 oTyrThrGluAlaAlaLeuValArgAspAspValAspTyrGlnIlePheA 38

     164 GCGACTTTGCCGAAAATAAAGGCAAGTTTGCAGTCGGGGCGAAAGATATT 213
         ||||||||||||||||||||||||||||:::|||||||||   :::::::
      38 rgAspPheAlaGluAsnLysGlyLysPheSerValGlyAlaThrAsnVal 54

     214 GAGGTTTACAACAAAAAAGGGGAGTTGGTCGGCAAATCGATGACGAAAGC 263
            |||   :::|||:::::::::   :::|||   ::::::   :::::
      55 LeuValLysAspLysAsnAsnLysAspLeuGlyThrAlaLeuProAsnGl 71

     264 C...CCGATGATTGATTTTTCTGTGGTATCG...CGTAACGGCGTGGCGG 307
         :   |||||||||||||||||||||||:::      :::   :::||||:
      71 yIleProMetIleAspPheSerValValAspValAspLysArgIleAlaT 88

     308 CATTGGCGGGCGATCAATATATTGTGAGCGTGGCACAT...AACGGCGGC 354
         ::|||   :::   ||||||:::|||:::|||   |||   ::::::|||
      88 hrLeuIleAsnProGlnTyrValValGlyValLysHisValSerAsnGly 104

     355 TATAACAATGTTGATTTTGGT..............GCGGAGGGAAGCAA 389
            :::::::::   ||||||                :::::::::||
     105 ValSerGluLeuHisPheGlyAsnLeuAsnGlyAsnMetAsnAsnGlyAs 121

     390 TCCCGATCAGCACCGC.................TTTTCTTACCAAATTG 421
         |         ||||||                        |||      |
     121 nAlaLysAlaHisArgAspValSerSerGluGluAsnArgTyrPheServ 138

     422 TGAAAAGAAATAATTATAAAGCAGGGACTAACGGC.............. 456
         ||:::::|||:::|||   :::      ||||||

     138 alGluLysAsnGluTyrProThrLysLeuAsnGlyLysThrValThrThr 154

     457 ...........CATCCTTATGGCGGCGATTATCATATGCCGCGTTTGCA 494
                     :::             ||||||:::||||||||||||
     155 GluAspGlnThrGlnLysArgArgGluAspTyrTyrMetProArgLeuAs 171

     495 CAAATTTGTCACAGATGCAGAACCTGTTGAG...ATGACCAGTTATATGG 541
         ||||||||||||:::::   |||:::|||      |||:::      |
     171 pLysPheValThrGluValAlaProIleGluAlaSerThrAlaSerSerA 188

     542 ATGGGTGGAAATACGCTGATTTAAATAAATACCCTGATCGTGTTCGAATC 591
         ||:::      |||   |||   ||||||||||||      ||||||:::
     188 spAlaGlyThrTyrAsnAspGlnAsnLysTyrProAlaPheValArgLeu 204

     592 GGAGCAGGCAGACAATATTGGCGGTCTGATGAAGACGAA........... 630
         |||:::|||   |||:::      :::      |||:::
     205 GlySerGlySerGlnPheIleTyrLysLysGlyAspAsnTyrSerLeuIl 221

     631 .CCCAATAACCGCGAAAGTTCATATCAT............ATTGCAAGCG 667
             ||||||:::|||   :::   :::            :::::::::|
     221 eLeuAsnAsnHisGluValGlyGlyAsnAsnLeuLysLeuValGlyAspA 238

     668 CATATTCTTGGCTCGTCGGTGGCAATACCTTTGCACAAAATGGATCAGGT 717
         |||||:::::   ::::::|||:::   :::      |||   ::::::
     238 laTyrThrTyrGlyIleAlaGlyThrProTyrLysValAsnHisGluAsn 254

     718 GGTGGCACAGTCAACTTAGGTAGCGAAAAAATTAAACATAGC...CCATA 764
         :::|||   :::::::::|||:::::|||   :::||||||   |||
     255 AsnGlyLeuIleGlyPheGlyAsnSerLysGluGluHisSerAspProLy 271

     765 TGGTTTTTTACCAACAGGA..............GGCTCATTTGGCGACA 799
          |||:::|||                  :::   :::|||||||
     271 sGlyIleLeuSerGlnAspProLeuThrAsnTyrAlaValLeuGlyAspS 288

     800 GTGGCTCACCAATGTTTATCTATGATGCCCAAAAGCAAAAGTGGTTAATT 849
         ||||||||||:::|||:::|||||||   :::|||   ||||||||||:::
     288 erGlySerProLeuPheValTyrAspArgGluLysGlyLysTrpLeuPhe 304

     850 AATGGGGTATTGCAAACAGGCAACCCCTATATAGGAAAAAGCAATGGCTT 899
            |||      :::            |||      ||||||
     305 LeuGlySerTyrAspPheTrpAlaGlyTyrAsnLysLysSer........ 318

     900 CCAGCTAGTTCGTAAAGATTGG...TTCTATGATGAAATCTTTGCTGGAG 946
                     :::::::|||   :::|||   :::   |||:::   |
     319 ..........TrpGlnGluTrpAsnIleTyrLysSerGlnPheThrLysA 332

     947 ATACCCATTCAGTATTCTACGAACCACATCAAAATGGGAAATACTTTTTT 996
         ||:::
     332 spValLeu..................................... 334

     997 AACGACAATAATAATGGCGCAGGAAAAATCGATGCCAAACATAAACACTA 1046
               |||:::::::::||||||::::::   ::::::          ||
     335 ......AsnLysAspSerAlaGlySerLeuIleGlySerLysThrAspTy 349

    1047 TTCTCTACCTTATAGATTAAAAACACGAACCGTT............CAAT 1084
         ||||         :::   ||||||  .|||:::            :::
     349 rSerTrpSerSerAsnGlyLysThrSerThrIleThrGlyGlyGluLysS 366

    1085 TGTTTAATGTTTCTTTATCCGAGACAGCAAGAGAACCTGTTTATCATGCT 1134
           :::||||||:::|||::::::          :::|||
     366 erLeuAsnValAspLeuAlaAspGlyLysAspLysPro............ 378

    1135 GCAGGTGGGGTCAACAGTTATCGACCCAGACTGAATAATGGAGAAAATAT 1184
                                            |||:::|||:::::::
     379 ............................AsnHisGlyLysSerVa 384

    1185 TTCCTTTATTGACAAAGGAAAAGGTGAATTGATACTTACCAGCAACATCA 1234
         ::::|||      :::|||:::|||   |||   |||::::::||||||:
     384 lThrPhe......GluGlySerGlyThrLeuThrLeuAsnAsnAsnIleA 399

    1235 ACCAAGGCGCGGGCGGTTTGTATTTTGAGGGTAATTTTACGGTC...TCG 1281
         ::|||||||||||||||||:::|||||||||::::::   |||   :::
     399 spGlnGlyAlaGlyGlyLeuPhePheGluGlyAspTyrGluValLysGly 415
```

```
1282 CCTAAAAACAACGAAACGTGGCAAGGCGCGGGCGTTCATATCAGTGATGG 1331
     :::::::|||   ||||||:::||||||||||||   ::::::::::||
 416 ThrSerAspAsnThrThrTrpLysGlyAlaGlyValSerValAlaGluGl 432

1332 CAGTACCGTTACTTGGAAAGTAAACGGCGTGGCAAACGACCGCCTGTCCA 1381
     |:::|||||||||||||||||||:::::          |||||||||:::|
 432 yLysThrValThrTrpLysValHisAsnProGlnTyrAspArgLeuAlaL 449

1382 AAATCGGCAAAGGCACGCTGCTGGTTCAAGCCAAAGGGGAAAACCAAGGC 1431
     ||||||||||||||||||||:::|||:::::   |||:::|||:::|||
 449 ysIleGlyLysGlyThrLeuIleValGluGlyThrGlyAspAsnLysGly 465

1432 TCGGTCAGCGTGGGCGACGGTAAAGTCATCTTAGATCAGCAGGCGGACGA 1481
     |||::::::|||||||||||||   |||||||||   ||||||::::::
 466 SerLeuLysValGlyAspGlyThrValIleLeuLysGlnGlnThrAsnGl 482

1482 TCAAGGCAAAAAACAAGCCTTTAGTGAAATCGGCTTGGTCAGCGGCAGGG 1531
      :::|||   ::::::||||||:::::::::|||:::||||||||||||:
 482 ySerGly...GlnHisAlaPheAlaSerValGlyIleValSerGlyArgS 498

1532 GGACGGTGCAACTGAATGCCGATAATCAGTTCAACCCCGACAAACTCTAT 1581
     ::|||:::   ||||||   |||:::|||   :::|||:::::::::|||
 498 erThrLeuValLeuAsnAspAspLysGlnValAspProAsnSerIleTyr 514

1582 TTCGGCTTTCGCGGCGGACGTTTGGATTTGAACGGGCATTCGCTTTCGTT 1631
     |||||||||||||||||||||||||||||||||||:::||||||:::||
 515 PheGlyPheArgGlyGlyArgLeuAspLeuAsnGlyAsnSerLeuThrPh 531

1632 CCACCGCATTCAAAATACCGATGAAGGGGCGATGATTGTCAACCACAATC 1681
     |   :::|||:::|||   |||:::||||||   :::||||||||||||:
 531 eAspHisIleArgAsnIleAspAspGlyAlaArgLeuValAsnHisAsnM 548

1682 AAGACAAAGAATCCACCGTTACCATTACAGGCAATAAAGATATTACT... 1728
     ::   :::   |||::::::||||||||||||:::::::   ||||||
 548 etThrAsnAlaSerAsnIleThrIleThrGlyGluSerLeuIleThrAsp 564

1728 ..................................................... 1728

 565 ProAsnThrIleThrProTyrAsnIleAspAlaProAspGluAspAsnPr 581

1728 ..................................................... 1728

 581 oTyrAlaPheArgArgIleLysAspGlyGlyGlnLeuTyrLeuAsnLeuG 598

1728 ..................................................... 1728

 598 luAsnTyrThrTyrTyrAlaLeuArgLysGlyAlaSerThrArgSerGlu 614

1729 .........ACAACCGGCAATAAC..........................  1743
              ::::::|||::::::
 615 LeuProLysAsnSerGlyGluSerAsnGluAsnTrpLeuTyrMetGlyLy 631

1744 ............................AACAACTTGGATAGCAAAA 1762
                                 |||:::::::::::::::::
 631 sThrSerAspGluAlaLysArgAsnValMetAsnHisIleAsnAsnGluA 648

1763 AAGAAATTGCCTACAACGGTTGGTTTGGCGAGAAAGATGCAACCAAAACG 1812
     ::      ::::::||||||:::|||||||||:::::::::   |||:::
 648 rgMetAsnGlyPheAsnGlyTyrPheGlyGluGluGluGly...LysAsn 663

1813 AACGGGCGGCTCAATCTGAATTACCAACCGGAAGAAGCGGATCGCACTTT 1862
     ||||||:::||||||:::::::::::   ::::::   ::::::   ::
 664 AsnGlyAsnLeuAsnValThrPheLysGlyLysSerGluGlnAsnArgPh 680

1863 ACTGCTTTCCGGCGGAACAAATTTAAACGGCAATATCACGCAAACAAACG 1912
     :||||||:::|||||||||||||||||||||||:::::|||      :::|
 680 eLeuLeuThrGlyGlyThrAsnLeuAsnGlyAspLeuThrValGluLysG 697

1913 GCAAACTGTTTTTCAGCGGCAGACCGACACCGCACGCCTACAATCATTTA 1962
     ||   ||||||:::|||||||||||||||||||||||||   :::
 697 lyThrLeuPheLeuSerGlyArgProThrProHisAlaArgAspIleAla 713
```

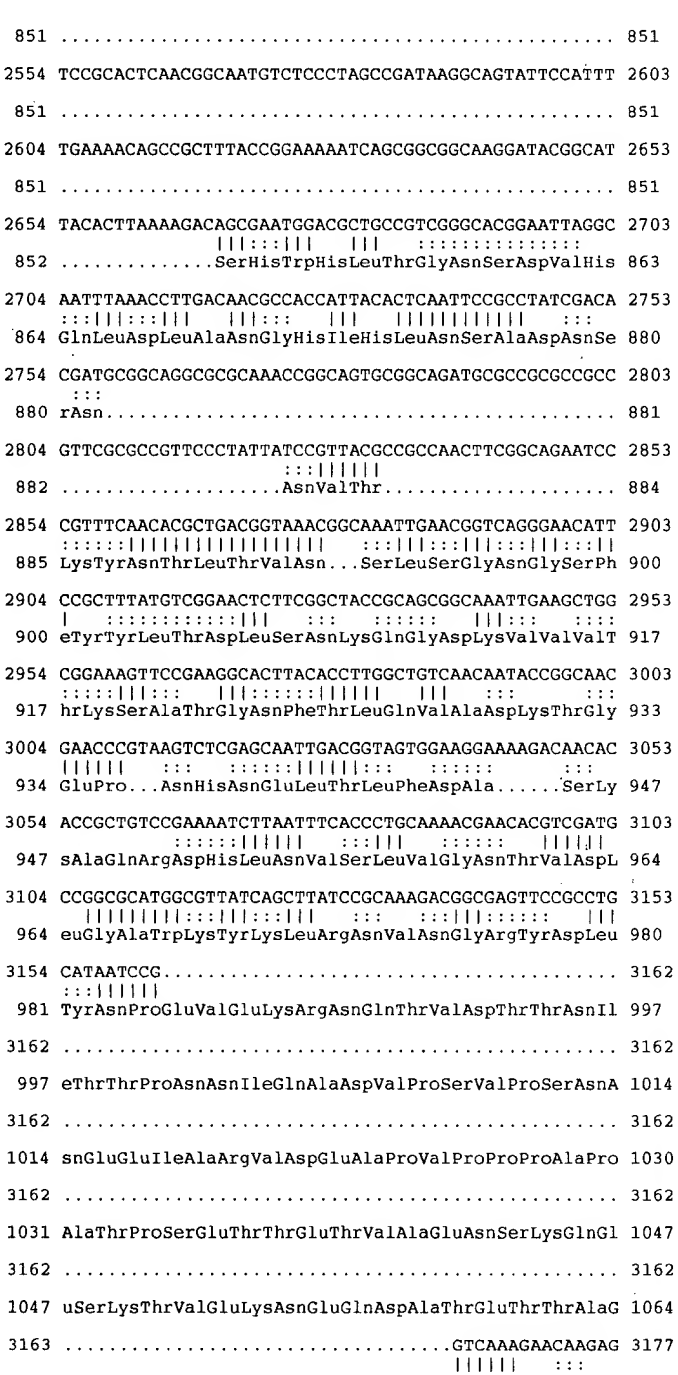
```
1963 GGAAGCGGGTGGTCAAAAATGGAAGGT......ATCCCACAAGGAGAAAT 2006
     |||   :::   :::|||   :::                 ::::::|||::
 714 GlyIleSerSerThrLysLysAspProHisPheAlaGluAsnAsnGluVa 730

2007 CGTGTGGGACAACGATTGGATCGACCGCACATTTAAAGCGGAAAACTTCC 2056
     :|||   ::::::|||||||||:::|||:::|||||||||   :::::::
 730 lValValGluAspAspTrpIleAsnArgAsnPheLysAlaThrThrMetA 747

2057 ATATTCAGGGCGGACAAGCGGTGGTTTCC...CGCAATGTTGCCAAAGTG 2103
     :::::   |||:::   ::::::   |||   ||||||||||||::::::
 747 snValThrGlyAsnAlaSerLeuTyrSerGlyArgAsnValAlaAsnIle 763

2104 GAAGGCGATTGGCATTTAAGCAATCACGCCCAAGCAGTTTTCGGTGTCGC 2153
        ::::::          ||||||   ||||||:::   :::|||
 764 ThrSerAsnIleThrAlaSerAsnLysAlaGlnValHisIleGlyTyr.. 779

2154 ACCGCATCAAAGCCACACAATCTGTACACGTTCGGACTGGACGGGTCTGA 2203
                ::::   |||:::|||:::||||||||||:::||||||   :
 780 ....LysThrGlyAspThrValCysValArgSerAspTyrThrGlyTyrV 795

2204 CAAGTTGTACCGAAAAAACCATTACCGACGATAAAGTGATTGCTTCATTG 2253
     :::::||||||          ::::::   ||||||::::::   |||:::
 795 alThrCysThrThrAspLysLeuSer...AspLysAlaLeuAsnSerPhe 810

2254 AGCAAGACCGACATCAGAGGCAATGTCAGCCTTGCCGATCACGCTCATTT 2303
     :::   |||:::::::||||||||||||:::|||::::::   |||:::::
 811 AsnProThrAsnLeuArgGlyAsnValAsnLeuThrGluSerAlaAsnPh 827

2304 AAATCTCACAGGACTTGCCACACTCAACGGCAATCTTAGTGCAGGCGGAG 2353
     :   |||   |||   |||:::|||   |||:::::::::::::   |||:
 827 eValLeu...GlyLysAlaAsnLeuPheGlyThrIleGlnSerArgGlyA 843

2354 ACACGCACTATACGGTTACGCGCAACGCCACCCAAAACGGCAACCTCAGC 2403
     ::::::::      :::|||:::|||
 843 snSerGlnValArgLeuThrGluAsn..................... 851

2404 CTCGTGGGCAATGCCCAAGCAACATTTAATCAAGCCACATTAAACGGCAA 2453

 851 ..................................................... 851

2454 CACATCGGCTTCGGACAATGCTTCATTTAATCTAAGCAACAACGCCGTAC 2503

 851 ..................................................... 851

2504 AAAACGGCAGTCTGACGCTTTCCGACAACGCTAAGGCAAACGTAAGCCAT 2553

 851 ..................................................... 851

2554 TCCGCACTCAACGGCAATGTCTCCCTAGCCGATAAGGCAGTATTCCATTT 2603

 851 ..................................................... 851

2604 TGAAAACAGCCGCTTTACCGGAAAAATCAGCGGCGGCAAGGATACGGCAT 2653

 851 ..................................................... 851

2654 TACACTTAAAAGACAGCGAATGGACGCTGCCGTCGGGCACGGAATTAGGC 2703
                  |||:::|||   |||   :::::::::::::::
 852 ..............SerHisTrpHisLeuThrGlyAsnSerAspValHis 863

2704 AATTTAAACCTTGACAACGCCACCATTACACTCAATTCCGCCTATCGACA 2753
     :::|||:::|||   |||:::   |||   ||||||||||||   :::
 864 GlnLeuAspLeuAlaAsnGlyHisIleHisLeuAsnSerAlaAspAsnSe 880

2754 CGATGCGGCAGGCGCGCAAACCGGCAGTGCGGCAGATGCGCCGCGCCGCC 2803
      :::
 880 rAsn..................................................  881

2804 GTTCGCGCCGTTCCCTATTATCCGTTACGCCGCCAACTTCGGCAGAATCC 2853
                          :::||||||
 882 ....................AsnValThr.....................  884

2854 CGTTTCAACACGCTGACGGTAAACGGCAAATTGAACGGTCAGGGAACATT 2903
```

```
          ::::::||||||||||||||||   :::|||:::|||:::|||:::||
 885 LysTyrAsnThrLeuThrValAsn...SerLeuSerGlyAsnGlySerPh 900

2904 CCGCTTTATGTCGGAACTCTTCGGCTACCGCAGCGGCAAATTGAAGCTGG 2953
     |  ::::::::::|||  :::  ::::::  |||:::  ::::
 900 eTyrTyrLeuThrAspLeuSerAsnLysGlnGlyAspLysValValValT 917

2954 CGGAAAGTTCCGAAGGCACTTACACCTTGGCTGTCAACAATACCGGCAAC 3003
     :::::|||:::   |||::::::||||||  |||   :::       :::
 917 hrLysSerAlaThrGlyAsnPheThrLeuGlnValAlaAspLysThrGly 933

3004 GAACCCGTAAGTCTCGAGCAATTGACGGTAGTGGAAGGAAAAGACAACAC 3053
     ||||||    :::   ::::::||||||:::   ::::::       :::
 934 GluPro...AsnHisAsnGluLeuThrLeuPheAspAla......SerLy 947

3054 ACCGCTGTCCGAAAATCTTAATTTCACCCTGCAAAACGAACACGTCGATG 3103
               ::::::||||||    :::|||   ::::::   ||||||
 947 sAlaGlnArgAspHisLeuAsnValSerLeuValGlyAsnThrValAspL 964

3104 CCGGCGCATGGCGTTATCAGCTTATCCGCAAAGACGGCGAGTTCCGCCTG 3153
        |||||||||:::|||:::|||    :::   :::|||::::::   |||
 964 euGlyAlaTrpLysTyrLysLeuArgAsnValAsnGlyArgTyrAspLeu 980

3154 CATAATCCG........................................ 3162
     :::||||||
 981 TyrAsnProGluValGluLysArgAsnGlnThrValAspThrThrAsnIl 997

3162 ................................................. 3162

 997 eThrThrProAsnAsnIleGlnAlaAspValProSerValProSerAsnA 1014

3162 ................................................. 3162

1014 snGluGluIleAlaArgValAspGluAlaProValProProProAlaPro 1030

3162 ................................................. 3162

1031 AlaThrProSerGluThrThrGluThrValAlaGluAsnSerLysGlnGl 1047

3162 ................................................. 3162

1047 uSerLysThrValGluLysAsnGluGlnAspAlaThrGluThrThrAlaG 1064

3163 ..................................GTCAAAGAACAAGAG 3177
                                       ||||||   :::
1064 lnAsnArgGluValAlaLysGluAlaLysSerAsnValLysAlaAsnThr 1080

3178 CTTTCCGACAAACTCGGCAAGGCGGGA...GAAACAGAGGCCGCCTTGAC 3224
        :::::::::::::::::::::|||   ||||||:::   :::   ||
1081 GlnThrAsnGluValAlaGlnSerGlySerGluThrLysGluThrGlnTh 1097

3225 GGCAAAACAGGCACAACTTGCCGCCAAACAACAGGCGGAAAAAGACAACG 3274
      |::::::     :::   |||:::   ::::::   ||||||   ::::
1097 rThrGluThrLysGluThrAlaThrValGluLysGluGluLysAlaLysV 1114

3275 CGCAAAGC.......................CTTGACGCGCTGATTGCG 3300
      ::::::::                        :::  :::   ::::::
1114 alGluThrGluLysThrGlnGluValProLysValThrSerGlnValSer 1130

3301 GCCGGGCGCAATGCCACCGAAAAGGCAGAAAGTGTTGCCGAACCGGCCCG 3350
              ::::::   :::|||   ::::::      |||||||||||||
1131 ProLysGlnGluGlnSerGluThrValGlnProGlnAlaGluProAlaAr 1147

3351 GCAGGCAGGCGGGGAAAATGCCGGCATTATGCAGGCGGAGGAAGAGAAAA 3400
      |             ||||||     ::::::   :::|||   ::::
1147 g...........GluAsnAspProThrValAsnIleLysGluProGlnS 1160

3401 AACGGGTGCAGGCGGATAAAGACACCGCCTTG...GCGAAACAGCGCGAA 3447
      ::::::::::::::        ||||||      |||||||:::   :::
1160 erGlnThrAsnThrThrAlaAspThrGluGlnProAlaLysGluThrSer 1176

3448 GCG.............................................. 3450
      :::

1177 SerAsnValGluGlnProValThrGluSerThrThrValAsnThrGlyAs 1193

3451 .....................GAAACCCGGCCGGCTACCACCGCCTTCC 3478
                          :::|||   |||||||||||       |
1193 nSerValValGluAsnProGluAsnThrThrProAlaThrThrGln...P 1209

3479 CC.................................CGCGCCCGCCGC 3492
     ||                                  |||   ||||||
1209 roThrValAsnSerGluSerSerAsnLysProLysAsnArgHisArgArg 1225

3493 GCCCGCCGGGATTTGCCG...CAACCGCAGCCCCAACCGCAACCCCAACC 3539
     :::   |||::::::|||   :::   :::|||            :::
1226 SerValArgSerValProHisAsnValGluProAlaThrThrSerSerAs 1242

3540 GCAGCGC.................GACCTGATCAGCCGTTATGCCAATA 3571
       :::|||                ||||||   |||       :::|||:
1242 nAspArgSerThrValAlaLeuCysAspLeuThrSerThrAsnThrAsnA 1259

3572 GCGGTTTGAGTGAATTTTCCGCC..............ACGCTCAACAGC 3606
     ::   ||||||:::          |||             :::||||||
1259 laValLeuSerAspAlaArgAlaLysAlaGlnPheValAlaLeuAsnVal 1275

3607 GTTTTCGCCGTACAGGACGAATTGGACCGCGTG...TTTGCCGAAGACCG 3653
          ||||||::::::::::::::::::::::   :::   ::::::
1276 GlyLysAlaValSerGlnHisIleSerGlnLeuGluMetAsnAsnGluGl 1292

3654 CCGCAACGCCGTTTGGACAAGCGGCATCCGGGACACCAAACACTACCGTT 3703
      :::       ||||||:::|||:::            :::|||:::|||   |
1292 yGlnTyrAsnValTrpValSerAsnThrSerMetAsnLysAsnTyrSerS 1309

3704 CGCAAGATTTCCGCGCCTACCGCCAACAAACCGACCTGCGCCAAATCGGT 3753
     ||::::::::::|||   :::   :::::::::           |||:::|||
1309 erSerGlnTyrArgArgPheSerSerLysSerThrGlnThrGlnLeuGly 1325

3754 ATGCAGAAAAACCTCGGCAGCGGG...CGCGTCGGCATCCTGTTTTCGCA 3800
        ::::::::::::::::::::::   ::::::|||   :::|||:::::
1326 TrpAspGlnThrIleSerAsnAsnValGlnLeuGlyGlyValPheThrTy 1342

3801 CAACCGGACCGGAAACACCTTCGACGACGGCATCGGCAACTCGGCACGGC 3850
     :   |||::::::|||:::||||||   :::   ::::::::::::
1342 rValArgAsnSerAsnAsnPheAspLysAlaThrSerLysAsnThrLeuA 1359

3851 TTGCCCACGGTGCCGTTTTCGGGCAATACGGCATCGGCAGGTTCGACATC 3900
              :::                |||:::   :::::::::   :::
1359 laGlnValAsnPheTyrSerLysTyrTyrAlaAspAsnHisTrpTyrLeu 1375

3901 GGCATCAGCGCGGGCGCGGGTTTTAGTAGCGGCAGCCTTTCAGACGGCAT 3950
     ||||||:::   |||   |||        :::::::::|||:::   :::
1376 GlyIleAspLeuGlyTyrGlyLysPheGlnSerLysLeuGlnThrAsnHi 1392

3951 CAGAGGCAAAATCCGCCGCCGCGTGCTGCATTACGGCATTCAGGCAAGAT 4000
        ::::::|||:::   |||::::::   ::::::|||:::   |||
1392 sAsnAlaLysPheAlaArgHisThrAlaGlnPheGlyLeuThrAlaGlyL 1409

4001 ACCGCGCAGGTTTCGGCGGATTCGGCATCGAACCGCACATCGGCGCAACG 4050
           ::::::|||:::||||||||||   |||   :::|||:::
1409 ysAlaPheAsnLeuGlyAsnPheGlyIleThrProIleValGlyValArg 1425

4051 CGCTATTTCGTCCAAAAAGCGGATTACCGATACGAAAACGTCAATATCGC 4100
          ::::::::::::||||||:::      ::::::::::::|||
1426 TyrSerTyrLeuSerAsnAlaAspPheAlaLeuAspGlnAlaArgIleLy 1442

4101 CACCCCGGGCCTTGCATTCAACCGCTACCGCGCGGGCATTAAGGCAGATT 4150
      :::      ::::::   :::          |||   :::      :::|
1442 sValAsnProIleSerValLysThrAlaPheAlaGlnValAspLeuSerT 1459

4151 ATTCATTCAAACCGGCGCAACACATTTCCATCACGCCTTATTTGAGCCTG 4200
     ||::::::         ::::::   :::|||:::||||||   ||||||
1459 yrThrTyrHisLeuGlyGlu...PheSerValThrProIleLeuSerAla 1474

4201 TCCTATACCGATGCCGCTTCCGGCAAAGTCCGAACGCGCGTCAATACCGC 4250
        |||   ||||||   :::|||:::   :::   :::||||||
1475 ArgTyr...AspAlaAsnGlnGlySerGlyLysIleAsnValAsnGlyTy 1490
```

```
     4251 CGTATTGGCGCAGGATTTCGGCAAAACCCGCAGTGCGGAATGGGGCGTAA 4300
     1490 rAspPheAlaTyrAsnValGluAsnGlnGlnGlnTyrAsnAlaGlyLeuL 1507

     4301 ACGCCGAAATCAAAGGTTTCACGCTGTCCCTCCACGCTGCCGCCGCCAAG 4350
     1507 ysLeuLysTyrHisAsnValLysLeuSerLeuIleGlyGlyLeuThrLys 1523

     4351 GGGCCGCAATTGGAAGCGCAGCACAGCGCGGGCATCAAATTAGGCTAC 4398
     1524 AlaLysGlnAlaGluLysGlnLysThrAlaGluLeuLysLeuSerPhe 1539

seq_name: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:PCT-US95-10661A-3

seq_documentation_block:
; Sequence 3, Application PC/TUS9510661A
;  GENERAL INFORMATION:
;    APPLICANT:  Washington University, et al.
;    TITLE OF INVENTION:  Haemophilus Adherence and Penetration Protein
;    NUMBER OF SEQUENCES:  9
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Flehr, Hohbach, Test, Albritton & Herbert
;      STREET:  4 Embarcadero Center, Suite 3400
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  United States
;      ZIP:  94111-4187
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US95/10661A
;      FILING DATE:  16-AUG-1995
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/296,791
;      FILING DATE:  25-AUG-1994
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Trecartin, Richard F.
;      REGISTRATION NUMBER:  31,801
;      REFERENCE/DOCKET NUMBER:  FP-59941/RFT
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 781-1989
;      TELEFAX:  (415) 398-3249
;      TELEX:  910 277299
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1541 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  unknown
PCT-US95-10661A-3


alignment_scores:
           Quality: 1614.00              Length:   1716
             Ratio:  1.732                 Gaps:     48
 Percent Similarity:  54.312   Percent Identity: 26.865

alignment_block:
 US-09-303-518D-653 x PCT-US95-10661A-3   ..

 Align seg 1/1  to: PCT-US95-10661A-3  from: 1  to: 1541

      64 CGCTTCTCGCCCGCTTACTTAGCCATATGCCTGTCGTTCGGCATTCTGCC 113
       5 LysPheLysLeuAsnPheIleAlaLeuThrValAlaTyrAlaLeuThrPr 21

     114 CCAAGCCCGGGCGGGACACACTTATTTCGGCATCAACTACCAATACTATC 163

      21 oTyrThrGluAlaAlaLeuValArgAspAspValAspTyrGlnIlePheA 38

     164 GCGACTTTGCCGAAAATAAAGGCAAGTTTGCAGTCGGGGCGAAAGATATT 213
      38 rgAspPheAlaGluAsnLysGlyLysPheSerValGlyAlaThrAsnVal 54

     214 GAGGTTTACAACAAAAAAGGGGAGTTGGTCGGCAAATCGATGACGAAAGC 263
      55 LeuValLysAspLysAsnAsnLysAspLeuGlyThrAlaLeuProAsnGl 71

     264 C...CCGATGATTGATTTTTCTGTGGTATCG...CGTAACGGCGTGGCGG 307
      71 yIleProMetIleAspPheSerValValAspValAspLysArgIleAlaT 88

     308 CATTGGCGGGCGATCAATATATTGTGAGCGTGGCACAT...AACGGCGGC 354
      88 hrLeuIleAsnProGlnTyrValValGlyValLysHisValSerAsnGly 104

     355 TATAACAATGTTGATTTTGGT..............GCGGAGGGAAGCAA 389
     105 ValSerGluLeuHisPheGlyAsnLeuAsnGlyAsnMetAsnAsnGlyAs 121

     390 TCCCGATCAGCACCGC.................TTTTCTTACCAAATTG 421
     121 nAlaLysAlaHisArgAspValSerSerGluGluAsnArgTyrPheSerV 138

     422 TGAAAAGAAATAATTATAAAGCAGGGACTAACGGC.............. 456
     138 alGluLysAsnGluTyrProThrLysLeuAsnGlyLysThrValThrThr 154

     457 ............CATCCTTATGGCGGCGATTATCATATGCCGCGTTTGCA 494
     155 GluAspGlnThrGlnLysArgArgGluAspTyrTyrMetProArgLeuAs 171

     495 CAAATTTGTCACAGATGCAGAACCTGTTGAG...ATGACCAGTTATATGG 541
     171 pLysPheValThrGluValAlaProIleGluAlaSerThrAlaSerSerA 188

     542 ATGGGTGGAAATACGCTGATTTAAATAAATACCCTGATCGTGTTCGAATC 591
     188 spAlaGlyThrTyrAsnAspGlnAsnLysTyrProAlaPheValArgLeu 204

     592 GGAGCAGGCAGACAATATTGGCGGTCTGATGAAGACGAA........... 630
     205 GlySerGlySerGlnPheIleTyrLysLysGlyAspAsnTyrSerLeuIl 221

     631 .CCCAATAACCGCGAAAGTTCATATCAT............ATTGCAAGCG 667
     221 eLeuAsnAsnHisGluValGlyGlyAsnAsnLeuLysLeuValGlyAspA 238

     668 CATATTCTTGGCTCGTCGGTGGCAATACCTTTGCACAAAATGGATCAGGT 717
     238 laTyrThrTyrGlyIleAlaGlyThrProTyrLysValAsnHisGluAsn 254

     718 GGTGGCACAGTCAACTTAGGTAGCGAAAAAATTAAACATAGC...CCATA 764
     255 AsnGlyLeuIleGlyPheGlyAsnSerLysGluGluHisSerAspProLy 271

     765 TGGTTTTTTACCAACAGGA..............GGCTCATTTGGCGACA 799
     271 sGlyIleLeuSerGlnAspProLeuThrAsnTyrAlaValLeuGlyAspS 288

     800 GTGGCTCACCAATGTTTATCTATGATGCCCAAAAGCAAAAGTGGTTAATT 849
     288 erGlySerProLeuPheValTyrAspArgGluLysGlyLysTrpLeuPhe 304

     850 AATGGGGTATTGCAAACAGGCAACCCCTATATAGGAAAAAGCAATGGCTT 899
     305 LeuGlySerTyrAspPheTrpAlaGlyTyrAsnLysLysSer........ 318

     900 CCAGCTAGTTCGTAAAGATTGG...TTCTATGATGAAATCTTTGCTGGAG 946
     319 .........TrpGlnGluTrpAsnIleTyrLysSerGlnPheThrLysA 332
```


947 ATACCCATTCAGTATTCTACGAACCACATCAAAATGGGAAATACTTTTTT 996
spValLeu 334
997 AACGACAATAATAATGGCGCAGGAAAAATCGATGCCAAACATAAACACTA 1046
AsnLysAspSerAlaGlySerLeuIleGlySerLysThrAspTy 349
1047 TTCTCTACCTTATAGATTAAAAACACGAACCGTT.....CAAT 1084
rSerTrpSerSerAsnGlyLysThrSerThrIleThrGlyGlyGluLysS 366
1085 TGTTTAATGTTTCTTTATCCGAGACAGCAAGAGAACCTGTTTATCATGCT 1134
erLeuAsnValAspLeuAlaAspGlyLysAspLysPro..... 378
1135 GCAGGTGGGGTCAACAGTTATCGACCCAGACTGAATAATGGAGAAAATAT 1184
AsnHisGlyLysServa 384
1185 TTCCTTTATTGACAAAGGAAAAGGTGAATTGATACTTACCAGCAACATCA 1234
lThrPhe.....GluGlySerGlyThrLeuThrLeuAsnAsnAsnIleA 399
1235 ACCAAGGCGCGGGCGGTTTGTATTTTGAGGGTAATTTTACGGTC...TCG 1281
spGlnGlyAlaGlyGlyLeuPhePheGluGlyAspTyrGluValLysGly 415
1282 CCTAAAAACAACGAAACGTGGCAAGGCGCGGGCGTTCATATCAGTGATGG 1331
ThrSerAspAsnThrThrTrpLysGlyAlaGlyValSerValAlaGluGl 432
1332 CAGTACCGTTACTTGGAAAGTAAACGGCGTGGCAAACGACCGCCTGTCCA 1381
yLysThrValThrTrpLysValHisAsnProGlnTyrAspArgLeuAlaL 449
1382 AAATCGGCAAAGGCACGCTGCTGGTTCAAGCCAAAGGGGAAAACCAAGGC 1431
ysIleGlyLysGlyThrLeuIleValGluGlyThrGlyAspAsnLysGly 465
1432 TCGGTCAGCGTGGGCGACGGTAAAGTCATCTTAGATCAGCAGGCGGACGA 1481
SerLeuLysValGlyAspGlyThrValIleLeuLysGlnGlnThrAsnGl 482
1482 TCAAGGCAAAAAACAAGCCTTTAGTGAAATCGGCTTGGTCAGCGGCAGGG 1531
ySerGly...GlnHisAlaPheAlaSerValGlyIleValSerGlyArgS 498
1532 GGACGGTGCAACTGAATGCCGATAATCAGTTCAACCCCGACAAACTCTAT 1581
erThrLeuValLeuAsnAspAspLysGlnValAspProAsnSerIleTyr 514
1582 TTCGGCTTTCGCGGCGGACGTTTGGATTTGAACGGGCATTCGCTTTCGTT 1631
PheGlyPheArgGlyGlyArgLeuAspLeuAsnGlyAsnSerLeuThrPh 531
1632 CCACCGCATTCAAAATACCGATGAAGGGGCGATGATTGTCAACCACAATC 1681
eAspHisIleArgAsnIleAspAspGlyAlaArgLeuValAsnHisAsnM 548
1682 AAGACAAAGAATCCACCGTTACCATTACAGGCAATAAAGATATTACT... 1728
etThrAsnAlaSerAsnIleThrIleThrGlyGluSerLeuIleThrAsp 564
1728 1728
ProAsnThrIleThrProTyrAsnIleAspAlaProAspGluAspAsnPr 581
1728 1728
oTyrAlaPheArgArgIleLysAspGlyGlyGlnLeuTyrLeuAsnLeuG 598


1728 1728
luAsnTyrThrTyrTyrAlaLeuArgLysGlyAlaSerThrArgSerGlu 614
1729 .....ACAACCGGCAATAAC..... 1743
LeuProLysAsnSerGlyGluSerAsnGluAsnTrpLeuTyrMetGlyLy 631
1744 .....AACAACTTGGATAGCAAAA 1762
sThrSerAspGluAlaLysArgAsnValMetAsnHisIleAsnAsnGluA 648
1763 AAGAAATTGCCTACAACGGTTGGTTTGGCGAGAAAGATGCAACCAAAACG 1812
rgMetAsnGlyPheAsnGlyTyrPheGlyGluGluGluGly...LysAsn 663
1813 AACGGGCGGCTCAATCTGAATTACCAACCGGAAGAAGCGGATCGCACTTT 1862
AsnGlyAsnLeuAsnValThrPheLysGlyLysSerGluGlnAsnArgPh 680
1863 ACTGCTTTCCGGCGGAACAAATTTAAACGGCAATATCACGCAAACAAACG 1912
eLeuLeuThrGlyGlyThrAsnLeuAsnGlyAspLeuThrValGluLysG 697
1913 GCAAACTGTTTTTCAGCGGCAGACCGACACCGCACGCCTACAATCATTTA 1962
lyThrLeuPheLeuSerGlyArgProThrProHisAlaArgAspIleAla 713
1963 GGAAGCGGGTGGTCAAAAATGGAAGGT......ATCCCACAAGGAGAAAT 2006
GlyIleSerSerThrLysLysAspProHisPheAlaGluAsnAsnGluVa 730
2007 CGTGTGGGACAACGATTGGATCGACCGCACATTTAAAGCGGAAAACTTCC 2056
lValValGluAspAspTrpIleAsnArgAsnPheLysAlaThrThrMetA 747
2057 ATATTCAGGGCGGACAAGCGGTGGTTTCC...CGCAATGTTGCCAAAGTG 2103
snValThrGlyAsnAlaSerLeuTyrSerGlyArgAsnValAlaAsnIle 763
2104 GAAGGCGATTGGCATTTAAGCAATCACGCCCAAGCAGTTTTCGGTGTCGC 2153
ThrSerAsnIleThrAlaSerAsnLysAlaGlnValHisIleGlyTyr.. 779
2154 ACCGCATCAAAGCCACACAATCTGTACACGTTCGGACTGGACGGGTCTGA 2203
....LysThrGlyAspThrValCysValArgSerAspTyrThrGlyTyrV 795
2204 CAAGTTGTACCGAAAAAACCATTACCGACGATAAAGTGATTGCTTCATTG 2253
alThrCysThrThrAspLysLeuSer...AspLysAlaLeuAsnSerPhe 810
2254 AGCAAGACCGACATCAGAGGCAATGTCAGCCTTGCCGATCACGCTCATTT 2303
AsnProThrAsnLeuArgGlyAsnValAsnLeuThrGluSerAlaAsnPh 827
2304 AAATCTCACAGGACTTGCCACACTCAACGGCAATCTTAGTGCAGGCGGAG 2353
eValLeu...GlyLysAlaAsnLeuPheGlyThrIleGlnSerArgGlyA 843
2354 ACACGCACTATACGGTTACGCGCAACGCCACCCAAAACGGCAACCTCAGC 2403
snSerGlnValArgLeuThrGluAsn..... 851
2404 CTCGTGGGCAATGCCCAAGCAACATTTAATCAAGCCACATTAAACGGCAA 2453
851 851
2454 CACATCGGCTTCGGACAATGCTTCATTTAATCTAAGCAACAACGCCGTAC 2503
851 851
2504 AAAACGGCAGTCTGACGCTTTCCGACAACGCTAAGGCAAACGTAAGCCAT 2553

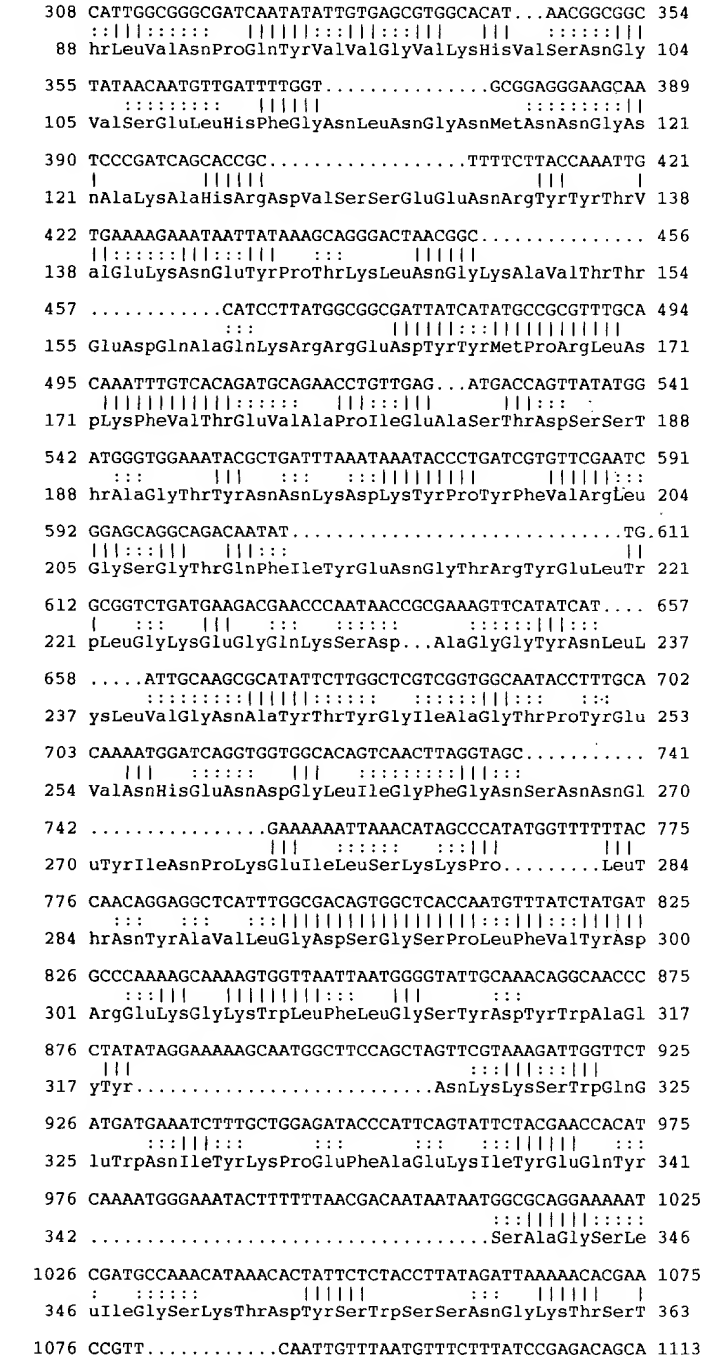

851 851
2554 TCCGCACTCAACGGCAATGTCTCCCTAGCCGATAAGGCAGTATTCCATTT 2603
851 851
2604 TGAAAACAGCCGCTTTACCGGAAAAATCAGCGGCGGCAAGGATACGGCAT 2653
851 851
2654 TACACTTAAAAGACAGCGAATGGACGCTGCCGTCGGGCACGGAATTAGGC 2703
852 .....SerHisTrpHisLeuThrGlyAsnSerAspValHis 863
2704 AATTTAAACCTTGACAACGCCACCATTACACTCAATTCCGCCTATCGACA 2753
864 GlnLeuAspLeuAlaAsnGlyHisIleHisLeuAsnSerAlaAspAsnSe 880
2754 CGATGCGGCAGGCGCGCAAACCGGCAGTGCGGCAGATGCGCCGCGCCGCC 2803
880 rAsn..... 881
2804 GTTCGCGCCGTTCCCTATTATCCGTTACGCCGCCAACTTCGGCAGAATCC 2853
882 .....AsnValThr..... 884
2854 CGTTTCAACACGCTGACGGTAAACGGCAAATTGAACGGTCAGGGAACATT 2903
885 LysTyrAsnThrLeuThrValAsn...SerLeuSerGlyAsnGlySerPh 900
2904 CCGCTTTATGTCGGAACTCTTCGGCTACCGCAGCGGCAAATTGAAGCTGG 2953
900 eTyrTyrLeuThrAspLeuSerAsnLysGlnGlyAspLysValValValT 917
2954 CGGAAAGTTCCGAAGGCACTTACACCTTGGCTGTCAACAATACCGGCAAC 3003
917 hrLysSerAlaThrGlyAsnPheThrLeuGlnValAlaAspLysThrGly 933
3004 GAACCCGTAAGTCTCGAGCAATTGACGGTAGTGGAAGGAAAAGACAACAC 3053
934 GluPro...AsnHisAsnGluLeuThrLeuPheAspAla.....SerLy 947
3054 ACCGCTGTCCGAAAATCTTAATTTCACCCTGCAAAACGAACACGTCGATG 3103
947 sAlaGlnArgAspHisLeuAsnValSerLeuValGlyAsnThrValAspL 964
3104 CCGGCGCATGGCGTTATCAGCTTATCCGCAAAGACGGCGAGTTCCGCCTG 3153
964 euGlyAlaTrpLysTyrLysLeuArgAsnValAsnGlyArgTyrAspLeu 980
3154 CATAATCCG..... 3162
981 TyrAsnProGluValGluLysArgAsnGlnThrValAspThrThrAsnIl 997
3162 3162
997 eThrThrProAsnAsnIleGlnAlaAspValProSerValProSerAsnA 1014
3162 3162
1014 snGluGluIleAlaArgValAspGluAlaProValProProProAlaPro 1030
3162 3162
1031 AlaThrProSerGluThrThrGluThrValAlaGluAsnSerLysGlnGl 1047
3162 3162
1047 uSerLysThrValGluLysAsnGluGlnAspAlaThrGluThrThrAlaG 1064
3163 .....GTCAAAGAACAAGAG 3177
1064 lnAsnArgGluValAlaLysGluAlaLysSerAsnValLysAlaAsnThr 1080
3178 CTTTCCGACAAACTCGGCAAGGCGGGA...GAAACAGAGGCCGCCTTGAC 3224
1081 GlnThrAsnGluValAlaGlnSerGlySerGluThrLysGluThrGlnTh 1097
3225 GGCAAAACAGGCACAACTTGCCGCCAAACAACAGGCGGAAAAAGACAACG 3274
1097 rThrGluThrLysGluThrAlaThrValGluLysGluGluLysAlaLysV 1114
3275 CGCAAAGC.....CTTGACGCGCTGATTGCG 3300
1114 alGluThrGluLysThrGlnGluValProLysValThrSerGlnValSer 1130
3301 GCCGGGCGCAATGCCACCGAAAAGGCAGAAAGTGTTGCCGAACCGGCCCG 3350
1131 ProLysGlnGluGlnSerGluThrValGlnProGlnAlaGluProAlaAr 1147
3351 GCAGGCAGGCGGGGAAAATGCCGGCATTATGCAGGCGGAGGAAGAGAAAA 3400
1147 g.....GluAsnAspProThrValAsnIleLysGluProGlnS 1160
3401 AACGGGTGCAGGCGGATAAAGACACCGCCTTG...GCGAAACAGCGCGAA 3447
1160 erGlnThrAsnThrThrAlaAspThrGluGlnProAlaLysGluThrSer 1176
3448 GCG..... 3450
1177 SerAsnValGluGlnProValThrGluSerThrThrValAsnThrGlyAs 1193
3451 .....GAAACCCGGCCGGCTACCACCGCCTTCC 3478
1193 nSerValValGluAsnProGluAsnThrThrProAlaThrThrGln...P 1209
3479 CC.....CGCGCCCGCCGC 3492
1209 roThrValAsnSerGluSerSerAsnLysProLysAsnArgHisArgArg 1225
3493 GCCCGCCGGGATTTGCCG...CAACCGCAGCCCCAACCGCAACCCCAACC 3539
1226 SerValArgSerValProHisAsnValGluProAlaThrThrSerSerAs 1242
3540 GCAGCGC.....GACCTGATCAGCCGTTATGCCAATA 3571
1242 nAspArgSerThrValAlaLeuCysAspLeuThrSerThrAsnThrAsnA 1259
3572 GCGGTTTGAGTGAATTTTCCGCC.....ACGCTCAACAGC 3606
1259 laValLeuSerAspAlaArgAlaLysAlaGlnPheValAlaLeuAsnVal 1275
3607 GTTTTCGCCGTACAGGACGAATTGGACCGCGTG...TTTGCCGAAGACCG 3653
1276 GlyLysAlaValSerGlnHisIleSerGlnLeuGluMetAsnAsnGluGl 1292
3654 CCGCAACGCCGTTTGGACAAGCGGCATCCGGGACACCAAACACTACCGTT 3703
1292 yGlnTyrAsnValTrpValSerAsnThrSerMetAsnLysAsnTyrSerS 1309
3704 CGCAAGATTTCCGCGCCTACCGCCAACAAACCGACCTGCGCCAAATCGGT 3753
1309 erSerGlnTyrArgArgPheSerSerLysSerThrGlnThrGlnLeuGly 1325
3754 ATGCAGAAAAACCTCGGCAGCGGG...CGCGTCGGCATCCTGTTTTCGCA 3800
1326 TrpAspGlnThrIleSerAsnAsnValGlnLeuGlyGlyValPheThrTy 1342
3801 CAACCGGACCGGAAACACCTTCGACGACGGCATCGGCAACTCGGCACGGC 3850
1342 rValArgAsnSerAsnAsnPheAspLysAlaThrSerLysAsnThrLeuA 1359
3851 TTGCCCACGGTGCCGTTTTCGGGCAATACGGCATCGGCAGGTTCGACATC 3900
1359 laGlnValAsnPheTyrSerLysTyrTyrAlaAspAsnHisTrpTyrLeu 1375

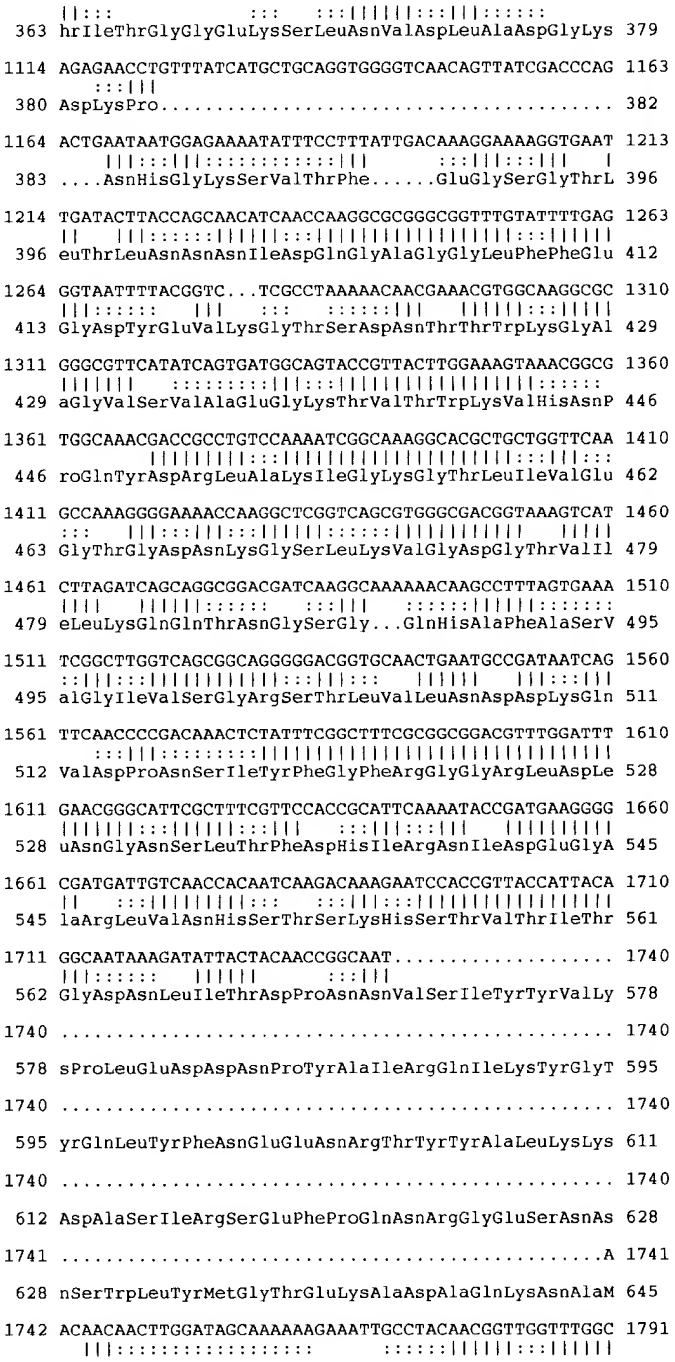
```
    3901 GGCATCAGCGCGGGCGCGGGTTTTAGTAGCGGCAGCCTTTCAGACGGCAT 3950
         ||||||:::   |||   |||      :::::::::|||:::   :::
    1376 GlyIleAspLeuGlyTyrGlyLysPheGlnSerLysLeuGlnThrAsnHi 1392

    3951 CAGAGGCAAAATCCGCCGCCGCGTGCTGCATTACGGCATTCAGGCAAGAT 4000
          :::::|||:::   |||:::::   ::::::|||:::   |||
    1392 sAsnAlaLysPheAlaArgHisThrAlaGlnPheGlyLeuThrAlaGlyL 1409

    4001 ACCGCGCAGGTTTCGGCGGATTCGGCATCGAACCGCACATCGGCGCAACG 4050
                 ::::::|||:::|||||||||   |||   :::|||:::
    1409 ysAlaPheAsnLeuGlyAsnPheGlyIleThrProIleValGlyValArg 1425

    4051 CGCTATTTCGTCCAAAAAGCGGATTACCGATACGAAAACGTCAATATCGC 4100
               ::::::::::::|||||||:::      ::::::::::::|||
    1426 TyrSerTyrLeuSerAsnAlaAspPheAlaLeuAspGlnAlaArgIleLy 1442

    4101 CACCCCGGGCCTTGCATTCAACCGCTACCGCGCGGGCATTAAGGCAGATT 4150
          :::      ::::::   :::         |||   :::      :::|
    1442 sValAsnProIleSerValLysThrAlaPheAlaGlnValAspLeuSerT 1459

    4151 ATTCATTCAAACCGGCGCAACACATTTCCATCACGCCTTATTTGAGCCTG 4200
         ||:::::::      ::::::   :::|||:::||||||   ||||||
    1459 yrThrTyrHisLeuGlyGlu...PheSerValThrProIleLeuSerAla 1474

    4201 TCCTATACCGATGCCGCTTCCGGCAAAGTCCGAACGCGCGTCAATACCGC 4250
            |||   ||||||   :::|||:::   :::   :::||||||
    1475 ArgTyr...AspAlaAsnGlnGlySerGlyLysIleAsnValAsnGlyTy 1490

    4251 CGTATTGGCGCAGGATTTCGGCAAAACCCGCAGTGCGGAATGGGGCGTAA 4300
          :::|||   :::      :::   ::::::   :::   |||::::
    1490 rAspPheAlaTyrAsnValGluAsnGlnGlnGlnTyrAsnAlaGlyLeuL 1507

    4301 ACGCCGAAATCAAAGGTTTCACGCTGTCCCTCCACGCTGCCGCCGCCAAG 4350
         ::   :::      :::      |||||||||   ::::::   :::|||
    1507 ysLeuLysTyrHisAsnValLysLeuSerLeuIleGlyGlyLeuThrLys 1523

    4351 GGGCCGCAATTGGAAGCGCAGCACAGCGCGGGCATCAAATTAGGCTAC 4398
         :::   |||   |||   |||   :::|||   :::||||||::::::
    1524 AlaLysGlnAlaGluLysGlnLysThrAlaGluLeuLysLeuSerPhe 1539

seq_name: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-08-296-791-4

seq_documentation_block:
; Sequence 4, Application US/08296791
; Patent No. 6245337
;  GENERAL INFORMATION:
;    APPLICANT:  St. Geme III, Joseph W.
;    APPLICANT:  Falkow, Stanley
;    TITLE OF INVENTION:  Haemophilus Adherence and Penetration
;    TITLE OF INVENTION:  Protein
;    NUMBER OF SEQUENCES:  9
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Flehr, Hohbach, Test, Albritton & Herbert
;      STREET:  4 Embarcadero Center, Suite 3400
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  United States
;      ZIP:  94111-4187
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/296,791
;      FILING DATE:  25-AUG-1994
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Trecartin, Richard F.
;      REGISTRATION NUMBER:  31,801
;      REFERENCE/DOCKET NUMBER:  A-59941/RFT/RMS
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 781-1989
;      TELEFAX:  (415) 398-3249
;      TELEX:  910 277299
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1545 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  unknown
US-08-296-791-4


alignment_scores:
            Quality: 1587.00              Length:   1723
              Ratio:  1.710                Gaps:     47
 Percent Similarity:  53.860   Percent Identity: 26.291

alignment_block:
 US-09-303-518D-653 x US-08-296-791-4   ..

 Align seg 1/1  to: US-08-296-791-4  from: 1  to: 1545

      64 CGCTTCTCGCCCGCTTACTTAGCCATATGCCTGTCGTTCGGCATTCTGCC 113
         :::|||:::      :::::|||:::   ::::::::::::::   ||
       5 LysPheLysLeuAsnPheIleAlaLeuThrValAlaTyrAlaLeuThrPr 21

     114 CCAAGCCCGGGCGGGACACACTTATTTCGGCATCAACTACCAATACTATC 163
         |   ::::::|||:::   :::         :::::|||||   :::|
      21 oTyrThrGluAlaAlaLeuValArgAspAspValAspTyrGlnIlePheA 38

     164 GCGACTTTGCCGAAAATAAAGGCAAGTTTGCAGTCGGGGCGAAAGATATT 213
         ||||||||||||||||||||||||||||:::|||||||||   ::::::
      38 rgAspPheAlaGluAsnLysGlyLysPheSerValGlyAlaThrAsnVal 54

     214 GAGGTTTACAACAAAAAAGGGGAGTTGGTCGGCAAATCGATGACGAAAGC 263
         ||||||   :::|||:::::::::   :::|||:::   :::   :::::
      55 GluValArgAspLysAsnAsnArgProLeuGlyAsnValLeuProAsnGl 71

     264 C...CCGATGATTGATTTTTCTGTGGTATCG...CGTAACGGCGTGGCGG 307
         :   |||||||||||||||||||||||:::      :::   :::|||:
      71 yIleProMetIleAspPheSerValValAspValAspLysArgIleAlaT 88

     308 CATTGGCGGGCGATCAATATATTGTGAGCGTGGCACAT...AACGGCGGC 354
         ::|||::::::   ||||||:::|||:::|||   |||   ::::::|||
      88 hrLeuValAsnProGlnTyrValValGlyValLysHisValSerAsnGly 104

     355 TATAACAATGTTGATTTTGGT..............GCGGAGGGAAGCAA 389
            :::::::::   ||||||              :::::::::||
     105 ValSerGluLeuHisPheGlyAsnLeuAsnGlyAsnMetAsnAsnGlyAs 121

     390 TCCCGATCAGCACCGC.................TTTTCTTACCAAATTG 421
         |         ||||||                        |||      |
     121 nAlaLysAlaHisArgAspValSerSerGluGluAsnArgTyrTyrThrV 138

     422 TGAAAAGAAATAATTATAAAGCAGGGACTAACGGC.............. 456
         ||:::::::|||:::|||   :::      ||||||
     138 alGluLysAsnGluTyrProThrLysLeuAsnGlyLysAlaValThrThr 154

     457 ............CATCCTTATGGCGGCGATTATCATATGCCGCGTTTGCA 494
                      :::            ||||||:::||||||||||||
     155 GluAspGlnAlaGlnLysArgArgGluAspTyrTyrMetProArgLeuAs 171

     495 CAAATTTGTCACAGATGCAGAACCTGTTGAG...ATGACCAGTTATATGG 541
          ||||||||||||::::::   |||:::|||      |||:::
     171 pLysPheValThrGluValAlaProIleGluAlaSerThrAspSerSerT 188

     542 ATGGGTGGAAATACGCTGATTTAAATAAATACCCTGATCGTGTTCGAATC 591
           :::      |||   :::   :::||||||||||      ||||||:::
     188 hrAlaGlyThrTyrAsnAsnLysAspLysTyrProTyrPheValArgLeu 204

     592 GGAGCAGGCAGACAATAT............................TG 611
         |||:::|||   |||:::                            ||
     205 GlySerGlyThrGlnPheIleTyrGluAsnGlyThrArgTyrGluLeuTr 221
```

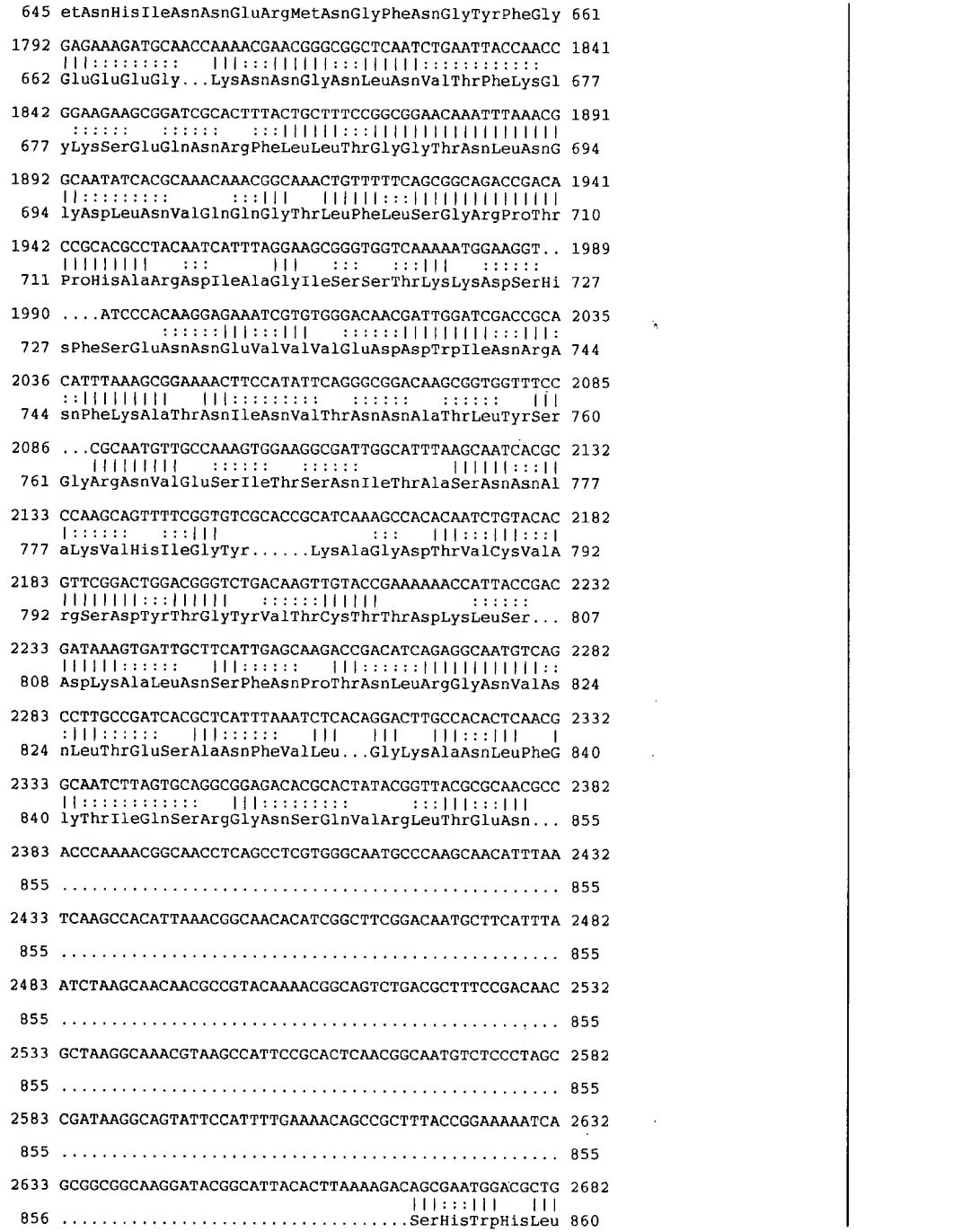
```
3551 TGATCAGCCGTTATGCCAATAGCGGTTTGAGTGAATTTTCCGCC..... 3594
1256 euThrSerThrAsnThrAsnAlaValLeuSerAspAlaArgAlaLysAla 1272

3595 .........ACGCTCAACAGCGTTTTCGCCGTACAGGACGAATTGGACCG 3635
1273 GlnPheValAlaLeuAsnValGlyLysAlaValSerGlnHisIleSerGl 1289

3636 CGTG...TTTGCCGAAGACCGCCGCAACGCCGTTTGGACAAGCGGCATCC 3682
1289 nLeuGluMetAsnAsnGluGlyGlnTyrAsnValTrpValSerAsnThrS 1306

3683 GGGACACCAAACACTACCGTTCGCAAGATTTCCGCGCCTACCGCCAACAA 3732
1306 erMetAsnLysAsnTyrSerSerSerGlnTyrArgArgPheSerSerLys 1322

3733 ACCGACCTGCGCCAAATCGGTATGCAGAAAAACCTCGGCAGCGGG...CG 3779
1323 SerThrGlnThrGlnLeuGlyTrpAspGlnThrIleSerAsnAsnValGl 1339

3780 CGTCGGCATCCTGTTTTCGCACAACCGGACCGGAAACACCTTCGACGACG 3829
1339 nLeuGlyGlyValPheThrTyrValArgAsnSerAsnAsnPheAspLysA 1356

3830 GCATCGGCAACTCGGCACGGCTTGCCCACGGTGCCGTTTTCGGGCAATAC 3879
1356 laThrSerLysAsnThrLeuAlaGlnValAsnPheTyrSerLysTyrTyr 1372

3880 GGCATCGGCAGGTTCGACATCGGCATCAGCGCGGGCGCGGGTTTTAGTAG 3929
1373 AlaAspAsnHisTrpTyrLeuGlyIleAspLeuGlyTyrGlyLysPheGl 1389

3930 CGGCAGCCTTTCAGACGGCATCAGAGGCAAAATCCGCCGCCGCGTGCTGC 3979
1389 nSerLysLeuGlnThrAsnHisAsnAlaLysPheAlaArgHisThrAlaG 1406

3980 ATTACGGCATTCAGGCAAGATACCGCGCAGGTTTCGGCGGATTCGGCATC 4029
1406 lnPheGlyLeuThrAlaGlyLysAlaPheAsnLeuGlyAsnPheGlyIle 1422

4030 GAACCGCACATCGGCGCAACGCGCTATTTCGTCCAAAAAGCGGATTACCG 4079
1423 ThrProIleValGlyValArgTyrSerTyrLeuSerAsnAlaAspPheAl 1439

4080 ATACGAAAACGTCAATATCGCCACCCCGGGCCTTGCATTCAACCGCTACC 4129
1439 aLeuAspGlnAlaArgIleLysValAsnProIleSerValLysThrAlaP 1456

4130 GCGCGGGCATTAAGGCAGATTATTCATTCAAACCGGCGCAACACATTTCC 4179
1456 heAlaGlnValAspLeuSerTyrThrTyrHisLeuGlyGlu...PheSer 1471

4180 ATCACGCCTTATTTGAGCCTGTCCTATACCGATGCCGCTTCCGGCAAAGT 4229
1472 ValThrProIleLeuSerAlaArgTyr...AspAlaAsnGlnGlySerGl 1487

4230 CCGAACGCGCGTCAATACCGCCGTATTGGCGCAGGATTTCGGCAAAACCC 4279
1487 yLysIleAsnValAsnGlyTyrAspPheAlaTyrAsnValGluAsnGlnG 1504

4280 GCAGTGCGGAATGGGGCGTAAACGCCGAAATCAAAGGTTTCACGCTGTCC 4329
1504 lnGlnTyrAsnAlaGlyLeuLysLeuLysTyrHisAsnValLysLeuSer 1520

4330 CTCCACGCTGCCGCCGCCAAGGGGCCGCAATTGGAAGCGCAGCACAGCGC 4379
1521 LeuIleGlyGlyLeuThrLysAlaLysGlnAlaGluLysGlnLysThrAl 1537

4380 GGGCATCAAATTAGGCTAC 4398
1537 aGluLeuLysLeuSerPhe 1543
```

seq_name: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:PCT-US95-10661A-4

```
seq_documentation_block:
; Sequence 4, Application PC/TUS9510661A
;  GENERAL INFORMATION:
;    APPLICANT:  Washington University, et al.
;    TITLE OF INVENTION:  Haemophilus Adherence and Penetration Protein
;    NUMBER OF SEQUENCES:  9
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Flehr, Hohbach, Test, Albritton & Herbert
;      STREET:  4 Embarcadero Center, Suite 3400
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  United States
;      ZIP:  94111-4187
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US95/10661A
;      FILING DATE:  16-AUG-1995
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/296,791
;      FILING DATE:  25-AUG-1994
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Trecartin, Richard F.
;      REGISTRATION NUMBER:  31,801
;      REFERENCE/DOCKET NUMBER:  FP-59941/RFT
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 781-1989
;      TELEFAX:  (415) 398-3249
;      TELEX:  910 277299
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1545 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  unknown
PCT-US95-10661A-4


alignment_scores:
           Quality: 1587.00              Length:   1723
             Ratio:  1.710                Gaps:     47
 Percent Similarity:  53.860  Percent Identity: 26.291

alignment_block:
 US-09-303-518D-653 x PCT-US95-10661A-4   ..

 Align seg 1/1  to: PCT-US95-10661A-4  from: 1  to: 1545

      64 CGCTTCTCGCCCGCTTACTTAGCCATATGCCTGTCGTTCGGCATTCTGCC 113
       5 LysPheLysLeuAsnPheIleAlaLeuThrValAlaTyrAlaLeuThrPr 21

     114 CCAAGCCCGGGCGGGACACACTTATTTCGGCATCAACTACCAATACTATC 163
      21 oTyrThrGluAlaAlaLeuValArgAspAspValAspTyrGlnIlePheA 38

     164 GCGACTTTGCCGAAAATAAAGGCAAGTTTGCAGTCGGGGCGAAAGATATT 213
      38 rgAspPheAlaGluAsnLysGlyLysPheSerValGlyAlaThrAsnVal 54

     214 GAGGTTTACAACAAAAAAGGGGAGTTGGTCGGCAAATCGATGACGAAAGC 263
      55 GluValArgAspLysAsnAsnArgProLeuGlyAsnValLeuProAsnGl 71

     264 C...CCGATGATTGATTTTTCTGTGGTATCG...CGTAACGGCGTGGCGG 307
      71 yIleProMetIleAspPheSerValValAspValAspLysArgIleAlaT 88
```






645 etAsnHisIleAsnAsnGluArgMetAsnGlyPheAsnGlyTyrPheGly 661
1792 GAGAAAGATGCAACCAAAACGAACGGGCGGCTCAATCTGAATTACCAACC 1841
662 GluGluGluGly...LysAsnAsnGlyAsnLeuAsnValThrPheLysGl 677
1842 GGAAGAAGCGGATCGCACTTTACTGCTTTCCGGCGGAACAAATTTAAACG 1891
677 yLysSerGluGlnAsnArgPheLeuLeuThrGlyGlyThrAsnLeuAsnG 694
1892 GCAATATCACGCAAACAAACGGCAAACTGTTTTTCAGCGGCAGACCGACA 1941
694 lyAspLeuAsnValGlnGlnGlyThrLeuPheLeuSerGlyArgProThr 710
1942 CCGCACGCCTACAATCATTTAGGAAGCGGGTGGTCAAAAATGGAAGGT.. 1989
711 ProHisAlaArgAspIleAlaGlyIleSerSerThrLysLysAspSerHi 727
1990 ....ATCCCACAAGGAGAAATCGTGTGGGACAACGATTGGATCGACCGCA 2035
727 sPheSerGluAsnAsnGluValValValGluAspAspTrpIleAsnArgA 744
2036 CATTTAAAGCGGAAAACTTCCATATTCAGGGCGGACAAGCGGTGGTTTCC 2085
744 snPheLysAlaThrAsnIleAsnValThrAsnAsnAlaThrLeuTyrSer 760
2086 ...CGCAATGTTGCCAAAGTGGAAGGCGATTGGCATTTAAGCAATCACGC 2132
761 GlyArgAsnValGluSerIleThrSerAsnIleThrAlaSerAsnAsnAl 777
2133 CCAAGCAGTTTTCGGTGTCGCACCGCATCAAAGCCACACAATCTGTACAC 2182
777 aLysValHisIleGlyTyr......LysAlaGlyAspThrValCysValA 792
2183 GTTCGGACTGGACGGGTCTGACAAGTTGTACCGAAAAAACCATTACCGAC 2232
792 rgSerAspTyrThrGlyTyrValThrCysThrThrAspLysLeuSer... 807
2233 GATAAAGTGATTGCTTCATTGAGCAAGACCGACATCAGAGGCAATGTCAG 2282
808 AspLysAlaLeuAsnSerPheAsnProThrAsnLeuArgGlyAsnValAs 824
2283 CCTTGCCGATCACGCTCATTTAAATCTCACAGGACTTGCCACACTCAACG 2332
824 nLeuThrGluSerAlaAsnPheValLeu...GlyLysAlaAsnLeuPheG 840
2333 GCAATCTTAGTGCAGGCGGAGACACGCACTATACGGTTACGCGCAACGCC 2382
840 lyThrIleGlnSerArgGlyAsnSerGlnValArgLeuThrGluAsn... 855
2383 ACCCAAAACGGCAACCTCAGCCTCGTGGGCAATGCCCAAGCAACATTTAA 2432
855 855
2433 TCAAGCCACATTAAACGGCAACACATCGGCTTCGGACAATGCTTCATTTA 2482
855 855
2483 ATCTAAGCAACAACGCCGTACAAAACGGCAGTCTGACGCTTTCCGACAAC 2532
855 855
2533 GCTAAGGCAAACGTAAGCCATTCCGCACTCAACGGCAATGTCTCCCTAGC 2582
855 855
2583 CGATAAGGCAGTATTCCATTTTGAAAACAGCCGCTTTACCGGAAAAATCA 2632
855 855
2633 GCGGCGGCAAGGATACGGCATTACACTTAAAAGACAGCGAATGGACGCTG 2682
856 .....SerHisTrpHisLeu 860
2683 CCGTCGGGCACGGAATTAGGCAATTTAAACCTTGACAACGCCACCATTAC 2732
861 ThrGlyAsnSerAspValHisGlnLeuAspLeuAlaAsnGlyHisIleHi 877
2733 ACTCAATTCCGCCTATCGACACGATGCGGCAGGCGCGCAAACCGGCAGTG 2782
877 sLeuAsnSerAlaAspAsnSerAsn..... 885
2783 CGGCAGATGCGCCGCGCCGCCGTTCGCGCCGTTCCCTATTATCCGTTACG 2832
886 .....AsnValThr 888
2833 CCGCCAACTTCGGCAGAATCCCGTTTCAACACGCTGACGGTAAACGGCAA 2882
889 .....LysTyrAsnThrLeuThrValAsn...Se 897
2883 ATTGAACGGTCAGGGAACATTCCGCTTTATGTCGGAACTCTTCGGCTACC 2932
897 rLeuSerGlyAsnGlySerPheTyrTyrLeuThrAspLeuSerAsnLysG 914
2933 GCAGCGGCAAATTGAAGCTGGCGGAAAGTTCCGAAGGCACTTACACCTTG 2982
914 lnGlyAspLysValValValThrLysSerAlaThrGlyAsnPheThrLeu 930
2983 GCTGTCAACAATACCGGCAACGAACCCGTAAGTCTCGAGCAATTGACGGT 3032
931 GlnValAlaAspLysThrGlyGluPro...AsnHisAsnGluLeuThrLe 946
3033 AGTGGAAGGAAAAGACAACACACCGCTGTCCGAAAATCTTAATTTCACCC 3082
946 uPheAspAla......SerLysAlaGlnArgAspHisLeuAsnValSerL 961
3083 TGCAAAACGAACACGTCGATGCCGGCGCATGGCGTTATCAGCTTATCCGC 3132
961 euValGlyAsnThrValAspLeuGlyAlaTrpLysTyrLysLeuArgAsn 977
3133 AAAGACGGCGAGTTCCGCCTGCATAATCCG..... 3162
978 ValAsnGlyArgTyrAspLeuTyrAsnProGluValGluLysArgAsnGl 994
3162 3162
994 nThrValAspThrThrAsnIleThrThrProAsnAsnIleGlnAlaAspV 1011
3162 3162
1011 alProSerValProSerAsnAsnGluGluIleAlaArgValAspGluAla 1027
3162 3162
1028 ProValProProProAlaProAlaThrProSerGluThrThrGluThrVa 1044
3162 3162
1044 lAlaGluAsnSerLysGlnGluSerLysThrValGluLysAsnGluGlnA 1061
3162 3162
1061 spAlaThrGluThrThrAlaGlnAsnArgGluValAlaLysGluAlaLys 1077
3163 ......GTCAAAGAACAAGAGCTTTCCGACAAACTCGGCAAGGCGGGA.. 3204
1078 SerAsnValLysAlaAsnThrGlnThrAsnGluValAlaGlnSerGlySe 1094
3205 .GAAACAGAGGCCGCCTTGACGGCAAAACAGGCACAACTTGCCGCCAAAC 3253
1094 rGluThrLysGluThrGlnThrThrGluThrLysGluThrAlaThrValG 1111
3254 AACAGGCGGAAAAAGACAACGCGCAAAGC..... 3282
1111 luLysGluGluLysAlaLysValGluThrGluLysThrGlnGluValPro 1127

```
3283 ...CTTGACGCGCTGATTGCGGCCGGGCGCAATGCCACCGAAAAGGCAGA 3329
1128 LysValThrSerGlnValSerProLysGlnGluGlnSerGluThrValGl 1144

3330 AAGTGTTGCCGAACCGGCCCGGCAGGCAGGCGGGGAAAATGCCGGCATTA 3379
1144 nProGlnAlaGluProAlaArg.....GluAsnAspProThrV 1157

3380 TGCAGGCGGAGGAAGAGAAAAAACGGGTGCAGGCGGATAAAGACACCGCC 3429
1157 alAsnIleLysGluProGlnSerGlnThrAsnThrThrAlaAspThrGlu 1173

3430 TTG...GCGAAACAGCGCGAAGCG..... 3450
1174 GlnProAlaLysGluThrSerSerAsnValGluGlnProValThrGluSe 1190

3451 .....GAAACCC 3457
1190 rThrThrValAsnThrGlyAsnSerValValGluAsnProGluAsnThrT 1207

3458 GGCCGGCTACCACCGCCTTCCCC..... 3480
1207 hrProAlaThrThrGln...ProThrValAsnSerGluSerSerAsnLys 1222

3481 .....CGCGCCCGCCGCGCCCGCCGGGATTTGCCG...CAACCGCA 3518
1223 ProLysAsnArgHisArgArgSerValArgSerValProHisAsnValGl 1239

3519 GCCCCAACCGCAACCCCAACCGCAGCGC.....GACC 3550
1239 uProAlaThrThrSerSerAsnAspArgSerThrValAlaLeuCysAspL 1256

3551 TGATCAGCCGTTATGCCAATAGCGGTTTGAGTGAATTTTCCGCC..... 3594
1256 euThrSerThrAsnThrAsnAlaValLeuSerAspAlaArgAlaLysAla 1272

3595 .....ACGCTCAACAGCGTTTTCGCCGTACAGGACGAATTGGACCG 3635
1273 GlnPheValAlaLeuAsnValGlyLysAlaValSerGlnHisIleSerGl 1289

3636 CGTG...TTTGCCGAAGACCGCCGCAACGCCGTTTGGACAAGCGGCATCC 3682
1289 nLeuGluMetAsnAsnGluGlyGlnTyrAsnValTrpValSerAsnThrS 1306

3683 GGGACACCAAACACTACCGTTCGCAAGATTTCCGCGCCTACCGCCAACAA 3732
1306 erMetAsnLysAsnTyrSerSerSerGlnTyrArgArgPheSerSerLys 1322

3733 ACCGACCTGCGCCAAATCGGTATGCAGAAAAACCTCGGCAGCGGG...CG 3779
1323 SerThrGlnThrGlnLeuGlyTrpAspGlnThrIleSerAsnAsnValGl 1339

3780 CGTCGGCATCCTGTTTTCGCACAACCGGACCGGAAACACCTTCGACGACG 3829
1339 nLeuGlyGlyValPheThrTyrValArgAsnSerAsnAsnPheAspLysA 1356

3830 GCATCGGCAACTCGGCACGGCTTGCCCACGGTGCCGTTTTCGGGCAATAC 3879
1356 laThrSerLysAsnThrLeuAlaGlnValAsnPheTyrSerLysTyrTyr 1372

3880 GGCATCGGCAGGTTCGACATCGGCATCAGCGCGGGCGCGGGTTTTAGTAG 3929
1373 AlaAspAsnHisTrpTyrLeuGlyIleAspLeuGlyTyrGlyLysPheGl 1389

3930 CGGCAGCCTTTCAGACGGCATCAGAGGCAAAATCCGCCGCCGCGTGCTGC 3979
1389 nSerLysLeuGlnThrAsnHisAsnAlaLysPheAlaArgHisThrAlaG 1406

3980 ATTACGGCATTCAGGCAAGATACCGCGCAGGTTTCGGCGGATTCGGCATC 4029
1406 lnPheGlyLeuThrAlaGlyLysAlaPheAsnLeuGlyAsnPheGlyIle 1422

4030 GAACCGCACATCGGCGCAACGCGCTATTTCGTCCAAAAAGCGGATTACCG 4079
1423 ThrProIleValGlyValArgTyrSerTyrLeuSerAsnAlaAspPheAl 1439

4080 ATACGAAAACGTCAATATCGCCACCCCGGGCCTTGCATTCAACCGCTACC 4129
1439 aLeuAspGlnAlaArgIleLysValAsnProIleSerValLysThrAlaP 1456

4130 GCGCGGGCATTAAGGCAGATTATTCATTCAAACCGGCGCAACACATTTCC 4179
1456 heAlaGlnValAspLeuSerTyrThrTyrHisLeuGlyGlu...PheSer 1471

4180 ATCACGCCTTATTTGAGCCTGTCCTATACCGATGCCGCTTCCGGCAAAGT 4229
1472 ValThrProIleLeuSerAlaArgTyr...AspAlaAsnGlnGlySerGl 1487

4230 CCGAACGCGCGTCAATACCGCCGTATTGGCGCAGGATTTCGGCAAAACCC 4279
1487 yLysIleAsnValAsnGlyTyrAspPheAlaTyrAsnValGluAsnGlnG 1504

4280 GCAGTGCGGAATGGGGCGTAAACGCCGAAATCAAAGGTTTCACGCTGTCC 4329
1504 lnGlnTyrAsnAlaGlyLeuLysLeuLysTyrHisAsnValLysLeuSer 1520

4330 CTCCACGCTGCCGCCGCCAAGGGGCCGCAATTGGAAGCGCAGCACAGCGC 4379
1521 LeuIleGlyGlyLeuThrLysAlaLysGlnAlaGluLysGlnLysThrAl 1537

4380 GGGCATCAAATTAGGCTAC 4398
1537 aGluLeuLysLeuSerPhe 1543
```

seq_name: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-08-296-791-5

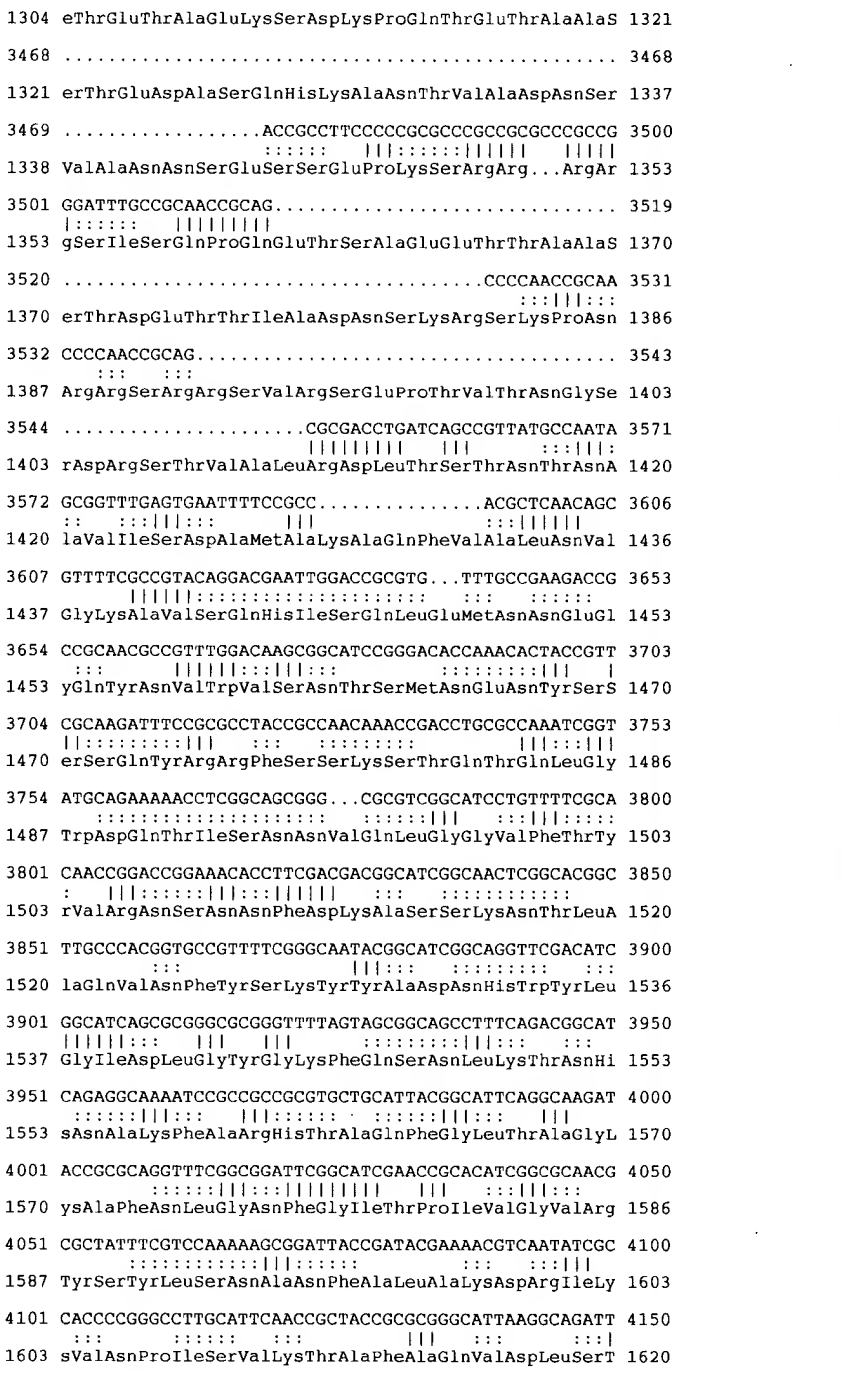
```
seq_documentation_block:
; Sequence 5, Application US/08296791
; Patent No. 6245337
;  GENERAL INFORMATION:
;    APPLICANT:  St. Geme III, Joseph W.
;    APPLICANT:  Falkow, Stanley
;    TITLE OF INVENTION:  Haemophilus Adherence and Penetration
;    TITLE OF INVENTION:  Protein
;    NUMBER OF SEQUENCES:  9
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Flehr, Hohbach, Test, Albritton & Herbert
;      STREET:  4 Embarcadero Center, Suite 3400
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  United States
;      ZIP:  94111-4187
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/296,791
;      FILING DATE:  25-AUG-1994
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Trecartin, Richard F.
;      REGISTRATION NUMBER:  31,801
;      REFERENCE/DOCKET NUMBER:  A-59941/RFT/RMS
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 781-1989
;      TELEFAX:  (415) 398-3249
;      TELEX:  910 277299
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1702 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  unknown
US-08-296-791-5
```

```
alignment_scores:
            Quality: 1581.50              Length:   1867
              Ratio:  1.677                 Gaps:     48
 Percent Similarity:  50.509    Percent Identity: 25.174

alignment_block:
 US-09-303-518D-653 x US-08-296-791-5   ..

 Align seg 1/1  to: US-08-296-791-5  from: 1  to: 1702

      64 CGCTTCTCGCCCGCTTACTTAGCCATATGCCTGTCGTTCGGCATTCTGCC 113
         :::|||:::      :::::::|||:::   ::::::::::::::   ||
       5 LysPheLysLeuAsnPheIleAlaLeuThrValAlaTyrAlaLeuThrPr 21

     114 CCAAGCCCGGGCGGGACACACTTATTTCGGCATCAACTACCAATACTATC 163
         |   :::::::|||:::   :::         ::::::||||||   :::|
      21 oTyrThrGluAlaAlaLeuValArgAspAspValAspTyrGlnIlePheA 38

     164 GCGACTTTGCCGAAAATAAAGGCAAGTTTGCAGTCGGGGCGAAAGATATT 213
         ||||||||||||||||||||||||:::|||:::||||||||   ::::::
      38 rgAspPheAlaGluAsnLysGlyArgPheSerValGlyAlaThrAsnVal 54

     214 GAGGTTTACAACAAAAAAGGGGAGTTGGTCGGCAAATCGATGACGAAAGC 263
         ||||||   :::|||:::::::::   :::|||:::   :::   :::::
      55 GluValArgAspLysAsnAsnHisSerLeuGlyAsnValLeuProAsnGl 71

     264 C...CCGATGATTGATTTTTCTGTGGTATCG...CGTAACGGCGTGGCGG 307
         :   |||||||||||||||||||||||:::      :::   :::|||:
      71 yIleProMetIleAspPheSerValValAspValAspLysArgIleAlaT 88

     308 CATTGGCGGGCGATCAATATATTGTGAGCGTGGCACAT...AACGGCGGC 354
         ::|||   :::   ||||||:::|||:::|||   |||   ::::::|||
      88 hrLeuIleAsnProGlnTyrValValGlyValLysHisValSerAsnGly 104

     355 TATAACAATGTTGATTTTGGT.............GCGGAGGGAAGCAA 389
            :::::::::   ||||||                :::::::::||
     105 ValSerGluLeuHisPheGlyAsnLeuAsnGlyAsnMetAsnAsnGlyAs 121

     390 TCCCGATCAGCACCGC.................TTTTCTTACCAAATTG 421
         |      :::||||||                     |||   .   |
     121 nAspLysSerHisArgAspValSerSerGluGluAsnArgTyrPheSerV 138

     422 TGAAAAGAAATAATTATAAAGCAGGGACTAACGGC.............. 456
         ||::::::|||:::|||   :::      ||||||
     138 alGluLysAsnGluTyrProThrLysLeuAsnGlyLysAlaValThrThr 154

     457 ............CATCCTTATGGCGGCGATTATCATATGCCGCGTTTGCA 494
                     :::            ||||||:::||||||||||||
     155 GluAspGlnThrGlnLysArgArgGluAspTyrTyrMetProArgLeuAs 171

     495 CAAATTTGTCACAGATGCAGAACCTGTTGAG...ATGACCAGTTATATGG 541
          ||||||||||||:::::   |||:::|||      |||:::      |
     171 pLysPheValThrGluValAlaProIleGluAlaSerThrAlaSerSerA 188

     542 ATGGGTGGAAATACGCTGATTTAAATAAATACCCTGATCGTGTTCGAATC 591
         ||:::      |||   |||   ||||||||||||      ||||||:::
     188 spAlaGlyThrTyrAsnAspGlnAsnLysTyrProAlaPheValArgLeu 204

     592 GGAGCAGGCAGACAATATTGGCGGTCTGATGAAGACGAA........... 630
         |||:::|||   |||:::      :::      |||:::
     205 GlySerGlyThrGlnPheIleTyrLysLysGlyAspAsnTyrSerLeuIl 221

     631 .CCCAATAACCGCGAAAGTTCATATCAT............ATTGCAAGCG 667
            ||||||:::|||   :::   :::           :::::::::|
     221 eLeuAsnAsnHisGluValGlyGlyAsnAsnLeuLysLeuValGlyAspA 238

     668 CATATTCTTGGCTCGTCGGTGGCAATACCTTTGCACAAAATGGATCAGGT 717
         |||||::::::   :::::::|||:::   :::      |||   ::::::
     238 laTyrThrTyrGlyIleAlaGlyThrProTyrLysValAsnHisGluAsn 254

     718 GGTGGCACAGTCAACTTAGGTAGCGAAAAAATTAAACATAGC...CCATA 764
         :::|||   :::::::::|||::::::|||   :::||||||   |||

     255 AsnGlyLeuIleGlyPheGlyAsnSerLysGluGluHisSerAspProLy 271

     765 TGGTTTTTTACCAACAGGA..............GGCTCATTTGGCGACA 799
           |||:::|||                     :::   :::|||||||
     271 sGlyIleLeuSerGlnAspProLeuThrAsnTyrAlaValLeuGlyAspS 288

     800 GTGGCTCACCAATGTTTATCTATGATGCCCAAAAGCAAAAGTGGTTAATT 849
         ||||||||||||:::|||:::||||||   :::|||   |||||||||:::
     288 erGlySerProLeuPheValTyrAspArgGluLysGlyLysTrpLeuPhe 304

     850 AATGGGGTATTGCAAACAGGCAACCCCTATATAGGAAAAAGCAATGGCTT 899
           |||      :::             |||      ||||||
     305 LeuGlySerTyrAspPheTrpAlaGlyTyrAsnLysLysSerTrpGlnGl 321

     900 CCAGCTAGTTCGTAAAGATTGGTTCTATGATGAAATCTTTGCTGGAGATA 949
               :::   |||      |||         ::::::      ||||
     321 uTrpAsnIleTyrLysProGluPheAlaLysThrValLeuAspLysAspT 338

     950 CCCATTCAGTATTCTACGAACCACATCAAAATGGGAAATACTTTTTTAAC 999
         ||
     338 hr..................................................  338

    1000 GACAATAATAATGGCGCAGGAAAAATCGATGCCAAACATAAACACTATTC 1049
                       ||||||::::::   :::::::::   :::|||::
     339 ..............AlaGlySerLeuIleGlySerAsnThrGlnTyrAs 350

    1050 TCTACCTTATAGATTAAAAACACGAACCGTT............CAATTGT 1087
         :                 ||||||   ::::::           :::   :
     350 nTrpAsnProThrGlyLysThrSerValIleSerAsnGlySerGluSerL 367

    1088 TTAATGTTTCTTTATCCGAGACAGCAAGAGAACCTGTTTATCATGCTGCA 1137
         ::|||||:::|||   ::::::::::::::
     367 euAsnValAspLeuPheAspSerSerGlnAsp..................  377

    1138 GGTGGGGTCAACAGTTATCGACCCAGACTGAATAATGGAGAAAATATTTC 1187
                    :::         :::   |||:::|||::::::::::::
     378 ............ThrAspSerLysLysAsnAsnHisGlyLysSerValTh 390

    1188 CTTTATTGACAAAGGAAAAGGTGAATTGATACTTACCAGCAACATCAACC 1237
         ::::      :::|||:::|||   |||   |||::::::||||||:::|
     390 rLeu......ArgGlySerGlyThrLeuThrLeuAsnAsnAsnIleAspG 405

    1238 AAGGCGCGGGCGGTTTGTATTTTGAGGGTAATTTTACGGTC...TCGCCT 1284
         ||||||||||||||||:::||||||||::::::   |||   :::
     405 lnGlyAlaGlyGlyLeuPhePheGluGlyAspTyrGluValLysGlyThr 421

    1285 AAAAACAACGAAACGTGGCAAGGCGCGGGCGTTCATATCAGTGATGGCAG 1334
         :::::::::   ||||||:::||||||||||||   ::::::||||||::
     422 SerAspSerThrThrTrpLysGlyAlaGlyValSerValAlaAspGlyLy 438

    1335 TACCGTTACTTGGAAAGTAAACGGCGTGGCAAACGACCGCCTGTCCAAAA 1384
         :|||||||||||||||||||:::::      :::||||||||||:::||||
     438 sThrValThrTrpLysValHisAsnProLysSerAspArgLeuAlaLysI 455

    1385 TCGGCAAAGGCACGCTGCTGGTTCAAGCCAAAGGGGAAAACCAAGGCTCG 1434
         ||||||||||||||||||:::|||:::::|||||||||||||:::||||||
     455 leGlyLysGlyThrLeuIleValGluGlyLysGlyGluAsnLysGlySer 471

    1435 GTCAGCGTGGGCGACGGTAAAGTCATCTTAGATCAGCAGGCGGACGATCA 1484
         ::::::||||||||||||   ||||||||   ||||||||||||   ::
     472 LeuLysValGlyAspGlyThrValIleLeuLysGlnGlnAlaAspAlaAs 488

    1485 AGGCAAAAAACAAGCCTTTAGTGAAATCGGCTTGGTCAGCGGCAGGGGGA 1534
         ::::|||   :::|||||||||:::::::|||:::||||||||||||:::|
     488 nAsnLysValLysAlaPheSerGlnValGlyIleValSerGlyArgSerT 505

    1535 CGGTGCAACTGAATGCCGATAATCAGTTCAACCCCGACAAACTCTATTTC 1584
         |||||   ||||||   |||:::|||   :::|||:::::::::||||||
     505 hrValValLeuAsnAspAspLysGlnValAspProAsnSerIleTyrPhe 521

    1585 GGCTTTCGCGGCGGACGTTTGGATTTGAACGGGCATTCGCTTTCGTTCCA 1634
         ||||||||||||||||||||||   ||||||:::::|||:::|||::
     522 GlyPheArgGlyGlyArgLeuAspAlaAsnGlyAsnAsnLeuThrPheGl 538
```

```
1635 CCGCATTCAAAATACCGATGAAGGGGCGATGATTGTCAACCACAATCAAG 1684
     ::::|||:::|||   |||:::||||||   :::||||||||||||   :
 538 uHisIleArgAsnIleAspAspGlyAlaArgLeuValAsnHisAsnThrS 555

1685 ACAAAGAATCCACCGTTACCATTACAGGCAATAAAGATATTACT..... 1728
     ::|||   ||||||||||||||||||||:::::   ||||||
 555 erLysThrSerThrValThrIleThrGlyGluSerLeuIleThrAspPro 571

1728 ..................................................... 1728

 572 AsnThrIleThrProTyrAsnIleAspAlaProAspGluAspAsnProTy 588

1728 ..................................................... 1728

 588 rAlaPheArgArgIleLysAspGlyGlyGlnLeuTyrLeuAsnLeuGluA 605

1728 ..................................................... 1728

 605 snTyrThrTyrTyrAlaLeuArgLysGlyAlaSerThrArgSerGluLeu 621

1729 ......ACAACCGGCAATAAC..................... 1743
           :::::::|||::::::
 622 ProLysAsnSerGlyGluSerAsnGluAsnTrpLeuTyrMetGlyLysTh 638

1744 ..........................AACAACTTGGATAGCAAAAAAG 1765
                                 |||:::::::::::::::::::
 638 rSerAspAlaAlaLysArgAsnValMetAsnHisIleAsnAsnGluArgM 655

1766 AAATTGCCTACAACGGTTGGTTTGGCGAGAAAGATGCAACCAAAACGAAC 1815
       ::::::||||||:::|||||||||:::::::::   |||:::|||
 655 etAsnGlyPheAsnGlyTyrPheGlyGluGluGluGly...LysAsnAsn 670

1816 GGGCGGCTCAATCTGAATTACCAACCGGAAGAAGCGGATCGCACTTTACT 1865
     |||:::||||||:::::::::::   ::::::   ::::::   :::||
 671 GlyAsnLeuAsnValThrPheLysGlyLysSerGluGlnAsnArgPheLe 687

1866 GCTTTCCGGCGGAACAAATTTAAACGGCAATATCACGCAAACAAACGGCA 1915
     ||||:::|||||||||||||||||||||:::::         :::|||
 687 uLeuThrGlyGlyThrAsnLeuAsnGlyAspLeuLysValGluLysGlyT 704

1916 AACTGTTTTTCAGCGGCAGACCGACACCGCACGCCTACAATCATTTAGGA 1965
       ||||||:::|||||||||||||||||||||||||   :::     |||
 704 hrLeuPheLeuSerGlyArgProThrProHisAlaArgAspIleAlaGly 720

1966 AGCGGGTGGTCAAAAATGGAAGGT......ATCCCACAAGGAGAAATCGT 2009
        :::   :::|||   :::               ::::::|||:::||
 721 IleSerSerThrLysLysAspGlnHisPheAlaGluAsnAsnGluValVa 737

2010 GTGGGACAACGATTGGATCGACCGCACATTTAAAGCGGAAAACTTCCATA 2059
     |   ::::::|||||||||:::|||:::|||||||||   |||:::::::
 737 lValGluAspAspTrpIleAsnArgAsnPheLysAlaThrAsnIleAsnV 754

2060 TTCAGGGCGGACAAGCGGTGGTTTCC...CGCAATGTTGCCAAAGTGGAA 2106
     ::   ::::::   ::::::   |||   ||||||||||||::::::
 754 alThrAsnAsnAlaThrLeuTyrSerGlyArgAsnValAlaAsnIleThr 770

2107 GGCGATTGGCATTTAAGCAATCACGCCCAAGCAGTTTTCGGTGTCGCACC 2156
     ::::::         |||:::::|||:::::   :::|||
 771 SerAsnIleThrAlaSerAspAsnAlaLysValHisIleGlyTyr..... 785

2157 GCATCAAAGCCACACAATCTGTACACGTTCGGACTGGACGGGTCTGACAA 2206
            :::   |||:::|||:::||||||||:::||||||   ::::
 786 .LysAlaGlyAspThrValCysValArgSerAspTyrThrGlyTyrValT 802

2207 GTTGTACCGAAAAAACCATTACCGACGATAAAGTGATTGCTTCATTGAGC 2256
     ::||||||         ::::::   ||||||:::::   |||:::::
 802 hrCysThrThrAspLysLeuSer...AspLysAlaLeuAsnSerPheAsn 817

2257 AAGACCGACATCAGAGGCAATGTCAGCCTTGCCGATCACGCTCATTTAAA 2306
        |||::::::   |||||||||:::|||:::
 818 AlaThrAsnValSerGlyAsnValAsnLeuSer..... 828
```

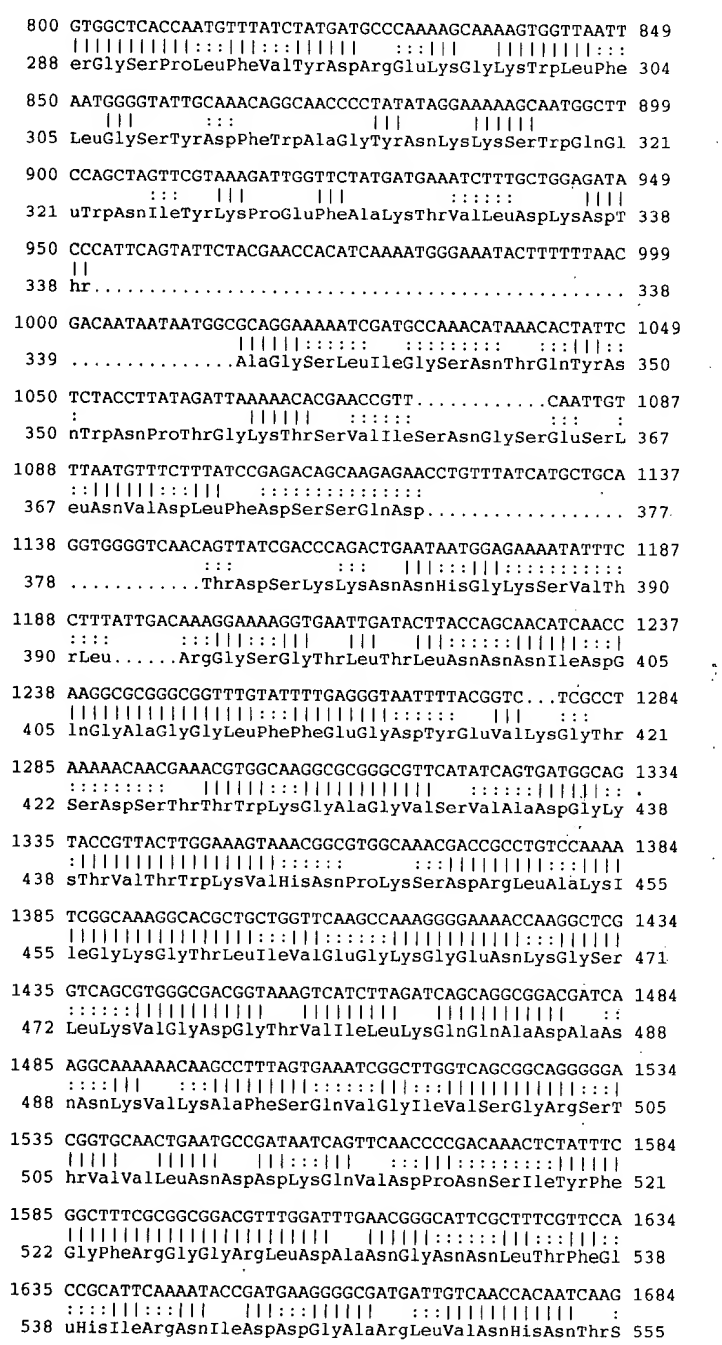
```
2307 TCTCACAGGACTTGCCACACTCAACGGCAATCTTAGTGCAGGCGGAGACA 2356

 828 ..................................................... 828

2357 CGCACTATACGGTTACGCGCAACGCCACCCAAAACGGCAACCTCAGCCTC 2406

 828 ..................................................... 828

2407 GTGGGCAATGCCCAAGCAACATTTAATCAAGCCACATTAAACGGCAACAC 2456
        ||||||||||:::   :::::::::::::|||:::|||   |||:::
 829 ...GlyAsnAlaAsnPheValLeuGlyLysAlaAsnLeuPheGlyThrIl 844

2457 ATCGGCTTCGGACAATGCTTCATTTAATCTAAGCAACAACGCCGTACAAA 2506
      |||:::::::   |||::::::   :::|||:::::|||
 844 eSerGlyThrGlyAsnSerGlnValArgLeuThrGluAsn.......... 857

2507 ACGGCAGTCTGACGCTTTCCGACAACGCTAAGGCAAACGTAAGCCATTCC 2556

 857 ..................................................... 857

2557 GCACTCAACGGCAATGTCTCCCTAGCCGATAAGGCAGTATTCCATTTTGA 2606

 857 ..................................................... 857

2607 AAACAGCCGCTTTACCGGAAAAATCAGCGGCGGCAAGGATACGGCATTAC 2656

 857 ..................................................... 857

2657 ACTTAAAAGACAGCGAATGGACGCTGCCGTCGGGCACGGAATTAGGCAAT 2706
                ||||:::|||   |||   :::   :::::::::::::::
 858 ...........SerHisTrpHisLeuThrGlyAspSerAsnValAsnGln 870

2707 TTAAACCTTGACAACGCCACCATTACACTCAATTCCGCCTATCGACACGA 2756
     ||||||||||||:::::   |||   ||||||:::       ::::::||
 871 LeuAsnLeuAspLysGlyHisIleHisLeuAsnAla......GlnAsnAs 885

2757 TGCGGCAGGCGCGCAAACCGGCAGTGCGGCAGATGCGCCGCGCCGCCGTT 2806
     ||||      :::   |||
 885 pAlaAsnLysValThrThr..... 891

2807 CGCGCCGTTCCCTATTATCCGTTACGCCGCCAACTTCGGCAGAATCCCGT 2856

 891 ..................................................... 891

2857 TTCAACACGCTGACGGTAAACGGCAAATTGAACGGTCAGGGAACATTCCG 2906
     :::|||||||||||||||||   :::|||:::|||:::|||:::|||
 892 TyrAsnThrLeuThrValAsn...SerLeuSerGlyAsnGlySerPheTy 907

2907 CTTTATGTCGGAACTCTTCGGCTACCGCAGCGGCAAATTGAAGCTGGCGG 2956
      :::::::::::|||   :::   ::::::   |||:::   :::::::
 907 rTyrLeuThrAspLeuSerAsnLysGlnGlyAspLysValValValThrL 924

2957 AAAGTTCCGAAGGCACTTACACCTTGGCTGTCAACAATACCGGCAACGAA 3006
     ::|||:::   |||::::::||||||   |||   :::       :::|||
 924 ysSerAlaThrGlyAsnPheThrLeuGlnValAlaAspLysThrGlyGlu 940

3007 CCCGTAAGTCTCGAGCAATTGACGGTAGTGGAAGGAAAAGACAACACACC 3056
     |||::::::   ::::::||||||:::   ::::::::::::   |||
 941 ProThrLys...AsnGluLeuThrLeuPheAspAlaSerAsnAlaThr.. 955

3057 GCTGTCCGAAAATCTTAATTTCACCCTGCAAAACGAACACGTCGATGCCG 3106
            :::|||||||||   :::|||   ::::::   ||||||   |
 956 ....ArgAsnAsnLeuAsnValSerLeuValGlyAsnThrValAspLeuG 971

3107 GCGCATGGCGTTATCAGCTTATCCGCAAAGACGGCGAGTTCCGCCTGCAT 3156
     ||||||||:::|||:::|||   :::   :::|||::::::   |||:::
 971 lyAlaTrpLysTyrLysLeuArgAsnValAsnGlyArgTyrAspLeuTyr 987

3157 AATCCG..... 3162
     ||||||
 988 AsnProGluValGluLysArgAsnGlnThrValAspThrThrAsnIleTh 1004

3162 ..................................................... 3162
```


1004 rThrProAsnAsnIleGlnAlaAspValProSerValProSerAsnAsnG 1021
3162 3162
1021 luGluIleAlaArgValGluThrProValProProProAlaProAlaThr 1037
3162 3162
1038 ProSerGluThrThrGluThrValAlaGluAsnSerLysGlnGluSerLy 1054
3163 .....GTCAAAGAACAAGAGCTTTCCGACAAACTCGGCAAGGCGG 3202
1054 sThrValGluLysAsnGluGlnAspAlaThrGluThrThrAlaGlnAsnG 1071
3203 GAGAA..... 3207
1071 lyGluValAlaGluGluAlaLysProSerValLysAlaAsnThrGlnThr 1087
3208 .....ACAGA 3212
1088 AsnGluValAlaGlnSerGlySerGluThrGluGluThrGlnThrThrGl 1104
3213 GGCCGCCTTGACGGCA..... 3228
1104 uIleLysGluThrAlaLysValGluLysGluGluLysAlaLysValGluL 1121
3228 3228
1121 ysGluGluLysAlaLysValGluLysAspGluIleGlnGluAlaProGln 1137
3229 .....AAACAGGCACAACTTGCCGCCAAA..... 3252
1138 MetAlaSerGluThrSerProLysGlnAlaLysProAlaProLysGluVa 1154
3252 3252
1154 lSerThrAspThrLysValGluGluThrGlnValGlnAlaGlnProGlnT 1171
3252 3252
1171 hrGlnSerThrThrValAlaAlaAlaGluAlaThrSerProAsnSerLys 1187
3253 .....CAACAGGCGGAAAAAGACAACGCGCAAAGCCTTGA 3287
1188 ProAlaGluGluThrGlnProSerGluLysThrAsnAlaGluProValTh 1204
3288 CGCGCTGATTGCGGCCGGGCGCAAT..... 3312
1204 rProValValSerLysAsnGlnThrGluAsnThrThrAspGlnProThrG 1221
3313 .....GCCACCGAAAAGGCAGAAAGTGTTGCC 3339
1221 luArgGluLysThrAlaLysValGluThrGluLysThrGlnGluProPro 1237
3340 GAACCGGCCCGGCAGGCAGGC..... 3360
1238 GlnValAlaSerGlnAlaSerProLysGlnGluGlnSerGluThrValGl 1254
3361 .....GGGGAAAATGCC...GGCATTATGCAGGCGG 3388
1254 nProGlnAlaValLeuGluSerGluAsnValProThrValAsnAsnAlaG 1271
3389 AGGAAGAGAAAAAACGGGTGCAGGCGGATAAAGAC..... 3423
1271 luGluValGlnAlaGlnLeuGlnThrGlnThrSerAlaThrValSerThr 1287
3424 .....ACCGCCTT 3431
1288 LysGlnProAlaProGluAsnSerIleAsnThrGlySerAlaThrAlaIl 1304
3432 GGCGAAACAGCGCGAAGCGGAAACCCGGCCGGCTACC..... 3468

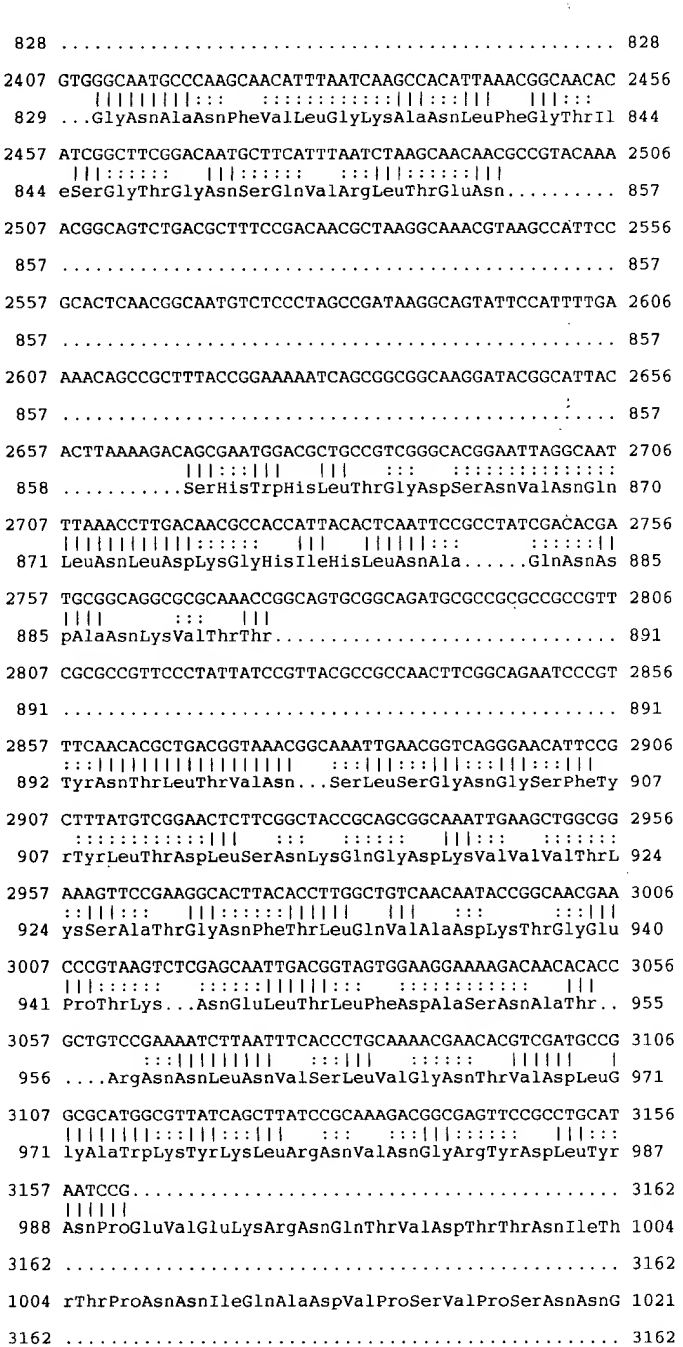

1304 eThrGluThrAlaGluLysSerAspLysProGlnThrGluThrAlaAlaS 1321
3468 3468
1321 erThrGluAspAlaSerGlnHisLysAlaAsnThrValAlaAspAsnSer 1337
3469 .....ACCGCCTTCCCCCGCGCCCGCCGCGCCCGCCG 3500
1338 ValAlaAsnAsnSerGluSerSerGluProLysSerArgArg...ArgAr 1353
3501 GGATTTGCCGCAACCGCAG..... 3519
1353 gSerIleSerGlnProGlnGluThrSerAlaGluGluThrThrAlaAlaS 1370
3520 .....CCCCAACCGCAA 3531
1370 erThrAspGluThrThrIleAlaAspAsnSerLysArgSerLysProAsn 1386
3532 CCCCAACCGCAG..... 3543
1387 ArgArgSerArgArgSerValArgSerGluProThrValThrAsnGlySe 1403
3544 .....CGCGACCTGATCAGCCGTTATGCCAATA 3571
1403 rAspArgSerThrValAlaLeuArgAspLeuThrSerThrAsnThrAsnA 1420
3572 GCGGTTTGAGTGAATTTTCCGCC.....ACGCTCAACAGC 3606
1420 laValIleSerAspAlaMetAlaLysAlaGlnPheValAlaLeuAsnVal 1436
3607 GTTTTCGCCGTACAGGACGAATTGGACCGCGTG...TTTGCCGAAGACCG 3653
1437 GlyLysAlaValSerGlnHisIleSerGlnLeuGluMetAsnAsnGluGl 1453
3654 CCGCAACGCCGTTTGGACAAGCGGCATCCGGGACACCAAACACTACCGTT 3703
1453 yGlnTyrAsnValTrpValSerAsnThrSerMetAsnGluAsnTyrSerS 1470
3704 CGCAAGATTTCCGCGCCTACCGCCAACAAACCGACCTGCGCCAAATCGGT 3753
1470 erSerGlnTyrArgArgPheSerSerLysSerThrGlnThrGlnLeuGly 1486
3754 ATGCAGAAAAACCTCGGCAGCGGG...CGCGTCGGCATCCTGTTTTCGCA 3800
1487 TrpAspGlnThrIleSerAsnAsnValGlnLeuGlyGlyValPheThrTy 1503
3801 CAACCGGACCGGAAACACCTTCGACGACGGCATCGGCAACTCGGCACGGC 3850
1503 rValArgAsnSerAsnAsnPheAspLysAlaSerSerLysAsnThrLeuA 1520
3851 TTGCCCACGGTGCCGTTTTCGGGCAATACGGCATCGGCAGGTTCGACATC 3900
1520 laGlnValAsnPheTyrSerLysTyrTyrAlaAspAsnHisTrpTyrLeu 1536
3901 GGCATCAGCGCGGGCGCGGGTTTTAGTAGCGGCAGCCTTTCAGACGGCAT 3950
1537 GlyIleAspLeuGlyTyrGlyLysPheGlnSerAsnLeuLysThrAsnHi 1553
3951 CAGAGGCAAAATCCGCCGCCGCGTGCTGCATTACGGCATTCAGGCAAGAT 4000
1553 sAsnAlaLysPheAlaArgHisThrAlaGlnPheGlyLeuThrAlaGlyL 1570
4001 ACCGCGCAGGTTTCGGCGGATTCGGCATCGAACCGCACATCGGCGCAACG 4050
1570 ysAlaPheAsnLeuGlyAsnPheGlyIleThrProIleValGlyValArg 1586
4051 CGCTATTTCGTCCAAAAAGCGGATTACCGATACGAAAACGTCAATATCGC 4100
1587 TyrSerTyrLeuSerAsnAlaAsnPheAlaLeuAlaLysAspArgIleLy 1603
4101 CACCCCGGGCCTTGCATTCAACCGCTACCGCGCGGGCATTAAGGCAGATT 4150
1603 sValAsnProIleSerValLysThrAlaPheAlaGlnValAspLeuSerT 1620

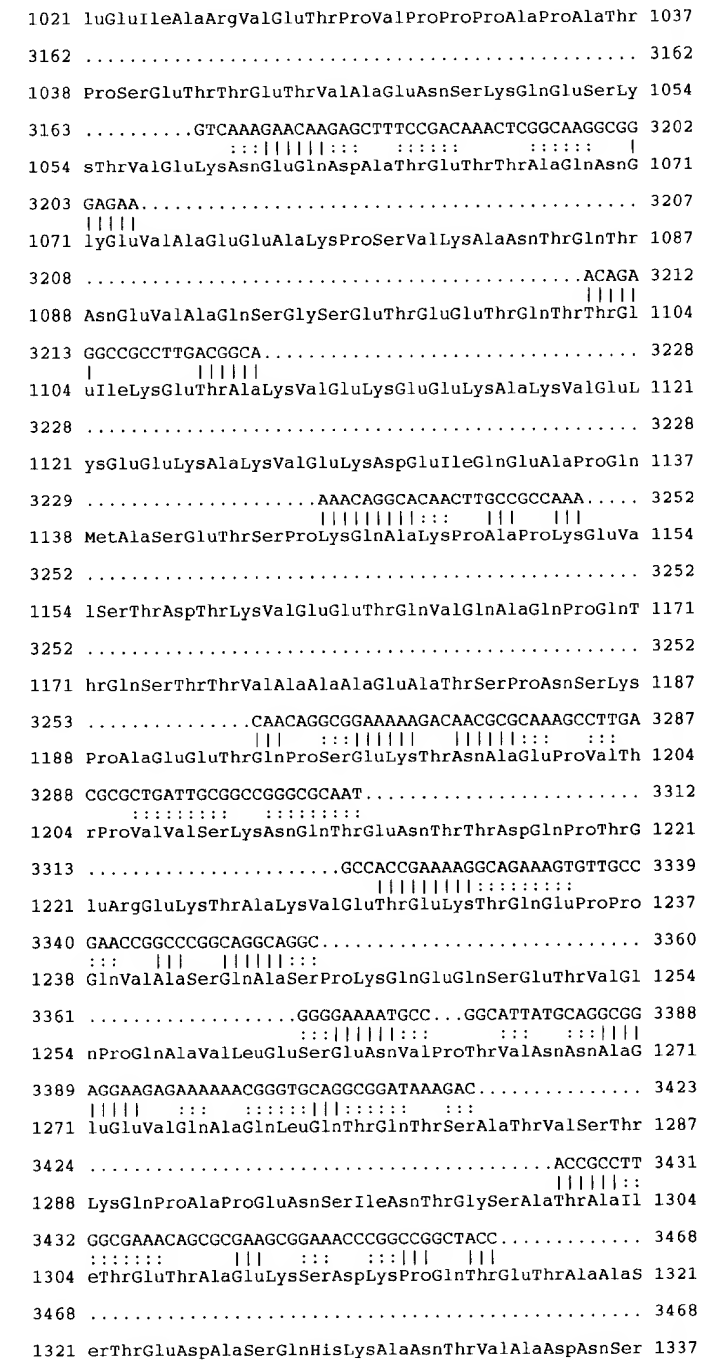
```
    4151 ATTCATTCAAACCGGCGCAACACATTTCCATCACGCCTTATTTGAGCCTG 4200
         ||:::::::      ::::::   :::|||:::||||||   ||||||
    1620 yrThrTyrHisLeuGlyGlu...PheSerValThrProIleLeuSerAla 1635

    4201 TCCTAT...ACCGATGCCGCTTCCGGCAAAGTCCGAACGCGCGTCAATAC 4247
           |||   |||:::   :::||||||||||:::      :::||||||
    1636 ArgTyrAspThrAsnGlnGlySerGlyLysIle......AsnValAsnGl 1650

    4248 CGCCGTATTGGCGCAGGATTTCGGCAAAACCCGCAGTGCGGAATGGGGCG 4297
              :::|||   :::      :::   ::::::   :::   |||:
    1650 nTyrAspPheAlaTyrAsnValGluAsnGlnGlnGlnTyrAsnAlaGlyL 1667

    4298 TAAACGCCGAAATCAAAGGTTTCACGCTGTCCCTCCACGCTGCCGCCGCC 4347
         :::::   :::      :::      |||||||||   ::::::   :::
    1667 euLysLeuLysTyrHisAsnValLysLeuSerLeuIleGlyGlyLeuThr 1683

    4348 AAGGGGCCGCAATTGGAAGCGCAGCACAGCGCGGGCATCAAATTAGGCTA 4397
         |||:::   |||   |||   |||   :::|||   :::||||||:::::
    1684 LysAlaLysGlnAlaGluLysGlnLysThrAlaGluLeuLysLeuSerPh 1700

    4398 C 4398
         :
    1700 e 1700

seq_name: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:PCT-US95-10661A-5

seq_documentation_block:
; Sequence 5, Application PC/TUS9510661A
;  GENERAL INFORMATION:
;    APPLICANT:  Washington University, et al.
;    TITLE OF INVENTION:  Haemophilus Adherence and Penetration Protein
;    NUMBER OF SEQUENCES:  9
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Flehr, Hohbach, Test, Albritton & Herbert
;      STREET:  4 Embarcadero Center, Suite 3400
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  United States
;      ZIP:  94111-4187
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US95/10661A
;      FILING DATE:  16-AUG-1995
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/296,791
;      FILING DATE:  25-AUG-1994
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Trecartin, Richard F.
;      REGISTRATION NUMBER:  31,801
;      REFERENCE/DOCKET NUMBER:  FP-59941/RFT
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 781-1989
;      TELEFAX:  (415) 398-3249
;      TELEX:  910 277299
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1702 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  unknown
PCT-US95-10661A-5


alignment_scores:
           Quality: 1581.50            Length:   1867
             Ratio:  1.677              Gaps:     48
 Percent Similarity:  50.509   Percent Identity: 25.174
```

```
alignment_block:
 US-09-303-518D-653 x PCT-US95-10661A-5   ..

 Align seg 1/1  to: PCT-US95-10661A-5  from: 1  to: 1702

     64 CGCTTCTCGCCCGCTTACTTAGCCATATGCCTGTCGTTCGGCATTCTGCC 113
        :::|||:::      ::::::|||:::   ::::::::::::::   ||
      5 LysPheLysLeuAsnPheIleAlaLeuThrValAlaTyrAlaLeuThrPr 21

    114 CCAAGCCCGGGCGGGACACACTTATTTCGGCATCAACTACCAATACTATC 163
        |   ::::::|||:::   :::         ::::::||||||   :::|
     21 oTyrThrGluAlaAlaLeuValArgAspAspValAspTyrGlnIlePheA 38

    164 GCGACTTTGCCGAAAATAAAGGCAAGTTTGCAGTCGGGGCGAAAGATATT 213
        ||||||||||||||||||||||:::|||:::|||||||||   ::::::
     38 rgAspPheAlaGluAsnLysGlyArgPheSerValGlyAlaThrAsnVal 54

    214 GAGGTTTACAACAAAAAAGGGGAGTTGGTCGGCAAATCGATGACGAAAGC 263
        ||||||   :::|||:::::::::   :::|||:::   :::   :::::
     55 GluValArgAspLysAsnAsnHisSerLeuGlyAsnValLeuProAsnGl 71

    264 C...CCGATGATTGATTTTTCTGTGGTATCG...CGTAACGGCGTGGCGG 307
        :   ||||||||||||||||||||||||:::      :::   :::|||:
     71 yIleProMetIleAspPheSerValValAspValAspLysArgIleAlaT 88

    308 CATTGGCGGGCGATCAATATATTGTGAGCGTGGCACAT...AACGGCGGC 354
        ::|||   :::   ||||||:::|||:::|||   |||   :::::|||
     88 hrLeuIleAsnProGlnTyrValValGlyValLysHisValSerAsnGly 104

    355 TATAACAATGTTGATTTTGGT..............GCGGAGGGAAGCAA 389
           :::::::   ||||||                 :::::::::||
    105 ValSerGluLeuHisPheGlyAsnLeuAsnGlyAsnMetAsnAsnGlyAs 121

    390 TCCCGATCAGCACCGC.................TTTTCTTACCAAATTG 421
        |      :::||||||                        |||      |
    121 nAspLysSerHisArgAspValSerSerGluGluAsnArgTyrPheServ 138

    422 TGAAAAGAAATAATTATAAAGCAGGGACTAACGGC.............. 456
        ||:::::::|||:::|||   :::      ||||||
    138 alGluLysAsnGluTyrProThrLysLeuAsnGlyLysAlaValThrThr 154

    457 ............CATCCTTATGGCGGCGATTATCATATGCCGCGTTTGCA 494
                       :::            ||||||:::|||||||||||||
    155 GluAspGlnThrGlnLysArgArgGluAspTyrTyrMetProArgLeuAs 171

    495 CAAATTTGTCACAGATGCAGAACCTGTTGAG...ATGACCAGTTATATGG 541
         ||||||||||||:::::   |||:::|||      |||:::      |
    171 pLysPheValThrGluValAlaProIleGluAlaSerThrAlaSerSerA 188

    542 ATGGGTGGAAATACGCTGATTTAAATAAATACCCTGATCGTGTTCGAATC 591
        ||:::      |||   |||   ||||||||||||      ||||||:::
    188 spAlaGlyThrTyrAsnAspGlnAsnLysTyrProAlaPheValArgLeu 204

    592 GGAGCAGGCAGACAATATTGGCGGTCTGATGAAGACGAA........... 630
        |||:::|||   |||:::      :::      |||:::
    205 GlySerGlyThrGlnPheIleTyrLysLysGlyAspAsnTyrSerLeuIl 221

    631 .CCCAATAACCGCGAAAGTTCATATCAT............ATTGCAAGCG 667
           ||||||:::|||   :::   :::            :::::::::|
    221 eLeuAsnAsnHisGluValGlyGlyAsnAsnLeuLysLeuValGlyAspA 238

    668 CATATTCTTGGCTCGTCGGTGGCAATACCTTTGCACAAAATGGATCAGGT 717
        ||||||::::::   ::::::|||:::   :::      |||   ::::::
    238 laTyrThrTyrGlyIleAlaGlyThrProTyrLysValAsnHisGluAsn 254

    718 GGTGGCACAGTCAACTTAGGTAGCGAAAAAATTAAACATAGC...CCATA 764
        :::|||   :::::::::|||::::::|||   :::||||||   |||
    255 AsnGlyLeuIleGlyPheGlyAsnSerLysGluGluHisSerAspProLy 271

    765 TGGTTTTTTACCAACAGGA..............GGCTCATTTGGCGACA 799
         |||:::|||                        :::   :::|||||||
    271 sGlyIleLeuSerGlnAspProLeuThrAsnTyrAlaValLeuGlyAspS 288
```


800 GTGGCTCACCAATGTTTATCTATGATGCCCAAAAGCAAAAGTGGTTAATT 849
288 erGlySerProLeuPheValTyrAspArgGluLysGlyLysTrpLeuPhe 304
850 AATGGGGTATTGCAAACAGGCAACCCCTATATAGGAAAAAGCAATGGCTT 899
305 LeuGlySerTyrAspPheTrpAlaGlyTyrAsnLysLysSerTrpGlnGl 321
900 CCAGCTAGTTCGTAAAGATTGGTTCTATGATGAAATCTTTGCTGGAGATA 949
321 uTrpAsnIleTyrLysProGluPheAlaLysThrValLeuAspLysAspT 338
950 CCCATTCAGTATTCTACGAACCACATCAAAATGGGAAATACTTTTTTAAC 999
338 hr..... 338
1000 GACAATAATAATGGCGCAGGAAAAATCGATGCCAAACATAAACACTATTC 1049
339 .....AlaGlySerLeuIleGlySerAsnThrGlnTyrAs 350
1050 TCTACCTTATAGATTAAAAACACGAACCGTT.....CAATTGT 1087
350 nTrpAsnProThrGlyLysThrSerValIleSerAsnGlySerGluSerL 367
1088 TTAATGTTTCTTTATCCGAGACAGCAAGAGAACCTGTTTATCATGCTGCA 1137
367 euAsnValAspLeuPheAspSerSerGlnAsp..... 377
1138 GGTGGGGTCAACAGTTATCGACCCAGACTGAATAATGGAGAAAATATTTC 1187
378 .....ThrAspSerLysLysAsnAsnHisGlyLysSerValTh 390
1188 CTTTATTGACAAAGGAAAAGGTGAATTGATACTTACCAGCAACATCAACC 1237
390 rLeu.....ArgGlySerGlyThrLeuThrLeuAsnAsnAsnIleAspG 405
1238 AAGGCGCGGGCGGTTTGTATTTTGAGGGTAATTTTACGGTC...TCGCCT 1284
405 lnGlyAlaGlyGlyLeuPhePheGluGlyAspTyrGluValLysGlyThr 421
1285 AAAAACAACGAAACGTGGCAAGGCGCGGGCGTTCATATCAGTGATGGCAG 1334
422 SerAspSerThrThrTrpLysGlyAlaGlyValSerValAlaAspGlyLy 438
1335 TACCGTTACTTGGAAAGTAAACGGCGTGGCAAACGACCGCCTGTCCAAAA 1384
438 sThrValThrTrpLysValHisAsnProLysSerAspArgLeuAlaLysI 455
1385 TCGGCAAAGGCACGCTGCTGGTTCAAGCCAAAGGGGAAAACCAAGGCTCG 1434
455 leGlyLysGlyThrLeuIleValGluGlyLysGlyGluAsnLysGlySer 471
1435 GTCAGCGTGGGCGACGGTAAAGTCATCTTAGATCAGCAGGCGGACGATCA 1484
472 LeuLysValGlyAspGlyThrValIleLeuLysGlnGlnAlaAspAlaAs 488
1485 AGGCAAAAAACAAGCCTTTAGTGAAATCGGCTTGGTCAGCGGCAGGGGGA 1534
488 nAsnLysValLysAlaPheSerGlnValGlyIleValSerGlyArgSerT 505
1535 CGGTGCAACTGAATGCCGATAATCAGTTCAACCCCGACAAACTCTATTTC 1584
505 hrValValLeuAsnAspAspLysGlnValAspProAsnSerIleTyrPhe 521
1585 GGCTTTCGCGGCGGACGTTTGGATTTGAACGGGCATTCGCTTTCGTTCCA 1634
522 GlyPheArgGlyGlyArgLeuAspAlaAsnGlyAsnAsnLeuThrPheGl 538
1635 CCGCATTCAAAATACCGATGAAGGGGCGATGATTGTCAACCACAATCAAG 1684
538 uHisIleArgAsnIleAspAspGlyAlaArgLeuValAsnHisAsnThrS 555


1685 ACAAAGAATCCACCGTTACCATTACAGGCAATAAAGATATTACT..... 1728
555 erLysThrSerThrValThrIleThrGlyGluSerLeuIleThrAspPro 571
1728 1728
572 AsnThrIleThrProTyrAsnIleAspAlaProAspGluAspAsnProTy 588
1728 1728
588 rAlaPheArgArgIleLysAspGlyGlyGlnLeuTyrLeuAsnLeuGluA 605
1728 1728
605 snTyrThrTyrTyrAlaLeuArgLysGlyAlaSerThrArgSerGluLeu 621
1729 ......ACAACCGGCAATAAC..... 1743
622 ProLysAsnSerGlyGluSerAsnGluAsnTrpLeuTyrMetGlyLysTh 638
1744 .....AACAACTTGGATAGCAAAAAAG 1765
638 rSerAspAlaAlaLysArgAsnValMetAsnHisIleAsnAsnGluArgM 655
1766 AAATTGCCTACAACGGTTGGTTTGGCGAGAAAGATGCAACCAAAACGAAC 1815
655 etAsnGlyPheAsnGlyTyrPheGlyGluGluGluGly...LysAsnAsn 670
1816 GGGCGGCTCAATCTGAATTACCAACCGGAAGAAGCGGATCGCACTTTACT 1865
671 GlyAsnLeuAsnValThrPheLysGlyLysSerGluGlnAsnArgPheLe 687
1866 GCTTTCCGGCGGAACAAATTTAAACGGCAATATCACGCAAACAAACGGCA 1915
687 uLeuThrGlyGlyThrAsnLeuAsnGlyAspLeuLysValGluLysGlyT 704
1916 AACTGTTTTTCAGCGGCAGACCGACACCGCACGCCTACAATCATTTAGGA 1965
704 hrLeuPheLeuSerGlyArgProThrProHisAlaArgAspIleAlaGly 720
1966 AGCGGGTGGTCAAAAATGGAAGGT......ATCCCACAAGGAGAAATCGT 2009
721 IleSerSerThrLysLysAspGlnHisPheAlaGluAsnAsnGluValVa 737
2010 GTGGGACAACGATTGGATCGACCGCACATTTAAAGCGGAAAACTTCCATA 2059
737 lValGluAspAspTrpIleAsnArgAsnPheLysAlaThrAsnIleAsnV 754
2060 TTCAGGGCGGACAAGCGGTGGTTTCC...CGCAATGTTGCCAAAGTGGAA 2106
754 alThrAsnAsnAlaThrLeuTyrSerGlyArgAsnValAlaAsnIleThr 770
2107 GGCGATTGGCATTTAAGCAATCACGCCCAAGCAGTTTTCGGTGTCGCACC 2156
771 SerAsnIleThrAlaSerAspAsnAlaLysValHisIleGlyTyr..... 785
2157 GCATCAAAGCCACACAATCTGTACACGTTCGGACTGGACGGGTCTGACAA 2206
786 .LysAlaGlyAspThrValCysValArgSerAspTyrThrGlyTyrValT 802
2207 GTTGTACCGAAAAAACCATTACCGACGATAAAGTGATTGCTTCATTGAGC 2256
802 hrCysThrThrAspLysLeuSer...AspLysAlaLeuAsnSerPheAsn 817
2257 AAGACCGACATCAGAGGCAATGTCAGCCTTGCCGATCACGCTCATTTAAA 2306
818 AlaThrAsnValSerGlyAsnValAsnLeuSer..... 828
2307 TCTCACAGGACTTGCCACACTCAACGGCAATCTTAGTGCAGGCGGAGACA 2356
828 828
2357 CGCACTATACGGTTACGCGCAACGCCACCCAAAACGGCAACCTCAGCCTC 2406


828 828
2407 GTGGGCAATGCCCAAGCAACATTTAATCAAGCCACATTAAACGGCAACAC 2456
829 ...GlyAsnAlaAsnPheValLeuGlyLysAlaAsnLeuPheGlyThrIl 844
2457 ATCGGCTTCGGACAATGCTTCATTTAATCTAAGCAACAACGCCGTACAAA 2506
844 eSerGlyThrGlyAsnSerGlnValArgLeuThrGluAsn..... 857
2507 ACGGCAGTCTGACGCTTTCCGACAACGCTAAGGCAAACGTAAGCCATTCC 2556
857 857
2557 GCACTCAACGGCAATGTCTCCCTAGCCGATAAGGCAGTATTCCATTTTGA 2606
857 857
2607 AAACAGCCGCTTTACCGGAAAAATCAGCGGCGGCAAGGATACGGCATTAC 2656
857 857
2657 ACTTAAAAGACAGCGAATGGACGCTGCCGTCGGGCACGGAATTAGGCAAT 2706
858 .....SerHisTrpHisLeuThrGlyAspSerAsnValAsnGln 870
2707 TTAAACCTTGACAACGCCACCATTACACTCAATTCCGCCTATCGACACGA 2756
871 LeuAsnLeuAspLysGlyHisIleHisLeuAsnAla.....GlnAsnAs 885
2757 TGCGGCAGGCGCGCAAACCGGCAGTGCGGCAGATGCGCCGCGCCGCCGTT 2806
885 pAlaAsnLysValThrThr..... 891
2807 CGCGCCGTTCCCTATTATCCGTTACGCCGCCAACTTCGGCAGAATCCCGT 2856
891 891
2857 TTCAACACGCTGACGGTAAACGGCAAATTGAACGGTCAGGGAACATTCCG 2906
892 TyrAsnThrLeuThrValAsn...SerLeuSerGlyAsnGlySerPheTy 907
2907 CTTTATGTCGGAACTCTTCGGCTACCGCAGCGGCAAATTGAAGCTGGCGG 2956
907 rTyrLeuThrAspLeuSerAsnLysGlnGlyAspLysValValValThrL 924
2957 AAAGTTCCGAAGGCACTTACACCTTGGCTGTCAACAATACCGGCAACGAA 3006
924 ysSerAlaThrGlyAsnPheThrLeuGlnValAlaAspLysThrGlyGlu 940
3007 CCCGTAAGTCTCGAGCAATTGACGGTAGTGGAAGGAAAAGACAACACACC 3056
941 ProThrLys...AsnGluLeuThrLeuPheAspAlaSerAsnAlaThr.. 955
3057 GCTGTCCGAAAATCTTAATTTCACCCTGCAAAACGAACACGTCGATGCCG 3106
956 ....ArgAsnAsnLeuAsnValSerLeuValGlyAsnThrValAspLeuG 971
3107 GCGCATGGCGTTATCAGCTTATCCGCAAAGACGGCGAGTTCCGCCTGCAT 3156
971 lyAlaTrpLysTyrLysLeuArgAsnValAsnGlyArgTyrAspLeuTyr 987
3157 AATCCG..... 3162
988 AsnProGluValGluLysArgAsnGlnThrValAspThrThrAsnIleTh 1004
3162 3162
1004 rThrProAsnAsnIleGlnAlaAspValProSerValProSerAsnAsnG 1021
3162 3162


1021 luGluIleAlaArgValGluThrProValProProProAlaProAlaThr 1037
3162 3162
1038 ProSerGluThrThrGluThrValAlaGluAsnSerLysGlnGluSerLy 1054
3163 .....GTCAAAGAACAAGAGCTTTCCGACAAACTCGGCAAGGCGG 3202
1054 sThrValGluLysAsnGluGlnAspAlaThrGluThrThrAlaGlnAsnG 1071
3203 GAGAA..... 3207
1071 lyGluValAlaGluGluAlaLysProSerValLysAlaAsnThrGlnThr 1087
3208 .....ACAGA 3212
1088 AsnGluValAlaGlnSerGlySerGluThrGluGluThrGlnThrThrGl 1104
3213 GGCCGCCTTGACGGCA..... 3228
1104 uIleLysGluThrAlaLysValGluLysGluGluLysAlaLysValGluL 1121
3228 3228
1121 ysGluGluLysAlaLysValGluLysAspGluIleGlnGluAlaProGln 1137
3229 .....AAACAGGCACAACTTGCCGCCAAA..... 3252
1138 MetAlaSerGluThrSerProLysGlnAlaLysProAlaProLysGluVa 1154
3252 3252
1154 lSerThrAspThrLysValGluGluThrGlnValGlnAlaGlnProGlnT 1171
3252 3252
1171 hrGlnSerThrThrValAlaAlaAlaGluAlaThrSerProAsnSerLys 1187
3253 .....CAACAGGCGGAAAAAGACAACGCGCAAAGCCTTGA 3287
1188 ProAlaGluGluThrGlnProSerGluLysThrAsnAlaGluProValTh 1204
3288 CGCGCTGATTGCGGCCGGGCGCAAT..... 3312
1204 rProValValSerLysAsnGlnThrGluAsnThrThrAspGlnProThrG 1221
3313 .....GCCACCGAAAAGGCAGAAAGTGTTGCC 3339
1221 luArgGluLysThrAlaLysValGluThrGluLysThrGlnGluProPro 1237
3340 GAACCGGCCCGGCAGGCAGGC..... 3360
1238 GlnValAlaSerGlnAlaSerProLysGlnGluGlnSerGluThrValGl 1254
3361 .....GGGGAAAATGCC...GGCATTATGCAGGCGG 3388
1254 nProGlnAlaValLeuGluSerGluAsnValProThrValAsnAsnAlaG 1271
3389 AGGAAGAGAAAAAACGGGTGCAGGCGGATAAAGAC..... 3423
1271 luGluValGlnAlaGlnLeuGlnThrGlnThrSerAlaThrValSerThr 1287
3424 .....ACCGCCTT 3431
1288 LysGlnProAlaProGluAsnSerIleAsnThrGlySerAlaThrAlaIl 1304
3432 GGCGAAACAGCGCGAAGCGGAAACCCGGCCGGCTACC..... 3468
1304 eThrGluThrAlaGluLysSerAspLysProGlnThrGluThrAlaAlaS 1321
3468 3468
1321 erThrGluAspAlaSerGlnHisLysAlaAsnThrValAlaAspAsnSer 1337

```
3469 ..................ACCGCCTTCCCCCGCGCCCGCCGCGCCCGCCG 3500
                       ::::::   |||::::::||||||   |||||
1338 ValAlaAsnAsnSerGluSerSerGluProLysSerArgArg...ArgAr 1353

3501 GGATTTGCCGCAACCGCAG............................... 3519
     |::::::   |||||||||
1353 gSerIleSerGlnProGlnGluThrSerAlaGluGluThrThrAlaAlaS 1370

3520 ......................................CCCCAACCGCAA 3531
                                           :::|||:::
1370 erThrAspGluThrThrIleAlaAspAsnSerLysArgSerLysProAsn 1386

3532 CCCCAACCGCAG...................................... 3543
       :::   :::                                 :
1387 ArgArgSerArgArgSerValArgSerGluProThrValThrAsnGlySe 1403

3544 ......................CGCGACCTGATCAGCCGTTATGCCAATA 3571
                           |||||||||   |||     :::|||:
1403 rAspArgSerThrValAlaLeuArgAspLeuThrSerThrAsnThrAsnA 1420

3572 GCGGTTTGAGTGAATTTTCCGCC...............ACGCTCAACAGC 3606
     ::   :::|||:::      |||               :::||||||
1420 laValIleSerAspAlaMetAlaLysAlaGlnPheValAlaLeuAsnVal 1436

3607 GTTTTCGCCGTACAGGACGAATTGGACCGCGTG...TTTGCCGAAGACCG 3653
          |||||::::::::::::::::::::   :::   ::::::
1437 GlyLysAlaValSerGlnHisIleSerGlnLeuGluMetAsnAsnGluGl 1453

3654 CCGCAACGCCGTTTGGACAAGCGGCATCCGGGACACCAAACACTACCGTT 3703
       :::      ||||||:::|||:::          :::::::::|||   |
1453 yGlnTyrAsnValTrpValSerAsnThrSerMetAsnGluAsnTyrSerS 1470

3704 CGCAAGATTTCCGCGCCTACCGCCAACAAACCGACCTGCGCCAAATCGGT 3753
     ||:::::::::|||   :::   :::::::::         |||:::|||
1470 erSerGlnTyrArgArgPheSerSerLysSerThrGlnThrGlnLeuGly 1486

3754 ATGCAGAAAAACCTCGGCAGCGGG...CGCGTCGGCATCCTGTTTTCGCA 3800
        ::::::::::::::::::::   ::::::|||   :::|||:::::
1487 TrpAspGlnThrIleSerAsnAsnValGlnLeuGlyGlyValPheThrTy 1503

3801 CAACCGGACCGGAAACACCTTCGACGACGGCATCGGCAACTCGGCACGGC 3850
     :   |||::::::|||:::||||||   :::   ::::::::::::
1503 rValArgAsnSerAsnAsnPheAspLysAlaSerSerLysAsnThrLeuA 1520

3851 TTGCCCACGGTGCCGTTTTCGGGCAATACGGCATCGGCAGGTTCGACATC 3900
        :::                         |||:::   ::::::::: :::
1520 laGlnValAsnPheTyrSerLysTyrTyrAlaAspAsnHisTrpTyrLeu 1536

3901 GGCATCAGCGCGGGCGCGGGTTTTAGTAGCGGCAGCCTTTCAGACGGCAT 3950
     |||||||:::   |||   |||       :::::::::|||:::   :::
1537 GlyIleAspLeuGlyTyrGlyLysPheGlnSerAsnLeuLysThrAsnHi 1553

3951 CAGAGGCAAAATCCGCCGCCGCGTGCTGCATTACGGCATTCAGGCAAGAT 4000
      ::::::|||:::   |||::::::   ::::::|||:::   |||
1553 sAsnAlaLysPheAlaArgHisThrAlaGlnPheGlyLeuThrAlaGlyL 1570

4001 ACCGCGCAGGTTTCGGCGGATTCGGCATCGAACCGCACATCGGCGCAACG 4050
                :::::::|||:::|||||||||   |||   :::|||:::
1570 ysAlaPheAsnLeuGlyAsnPheGlyIleThrProIleValGlyValArg 1586

4051 CGCTATTTCGTCCAAAAAGCGGATTACCGATACGAAAACGTCAATATCGC 4100
          :::::::::::|||::::::           :::   :::|||
1587 TyrSerTyrLeuSerAsnAlaAsnPheAlaLeuAlaLysAspArgIleLy 1603

4101 CACCCCGGGCCTTGCATTCAACCGCTACCGCGCGGGCATTAAGGCAGATT 4150
      :::      ::::::   :::            |||   :::     :::|
1603 sValAsnProIleSerValLysThrAlaPheAlaGlnValAspLeuSerT 1620

4151 ATTCATTCAAACCGGCGCAACACATTTCCATCACGCCTTATTTGAGCCTG 4200
     ||::::::         ::::::   :::|||:::||||||   ||||||
1620 yrThrTyrHisLeuGlyGlu...PheSerValThrProIleLeuSerAla 1635

4201 TCCTAT...ACCGATGCCGCTTCCGGCAAAGTCCGAACGCGCGTCAATAC 4247
        |||   |||:::   :::||||||||||:::      :::||||||
1636 ArgTyrAspThrAsnGlnGlySerGlyLysIle......AsnValAsnGl 1650

4248 CGCCGTATTGGCGCAGGATTTCGGCAAAACCCGCAGTGCGGAATGGGGCG 4297
          :::|||   :::         :::   ::::::   :::   |||:
1650 nTyrAspPheAlaTyrAsnValGluAsnGlnGlnGlnTyrAsnAlaGlyL 1667

4298 TAAACGCCGAAATCAAAGGTTTCACGCTGTCCCTCCACGCTGCCGCCGCC 4347
     :::::   :::      :::      |||||||||   ::::::   :::
1667 euLysLeuLysTyrHisAsnValLysLeuSerLeuIleGlyGlyLeuThr 1683

4348 AAGGGGCCGCAATTGGAAGCGCAGCACAGCGCGGGCATCAAATTAGGCTA 4397
     |||:::   |||   |||   |||   :::|||   :::||||||:::::
1684 LysAlaLysGlnAlaGluLysGlnLysThrAlaGluLeuLysLeuSerPh 1700

4398 C 4398
     :
1700 e 1700
```

seq_name: /cgn2_6/ptodata/1/iaa/backfiles1.pep:5268270-2

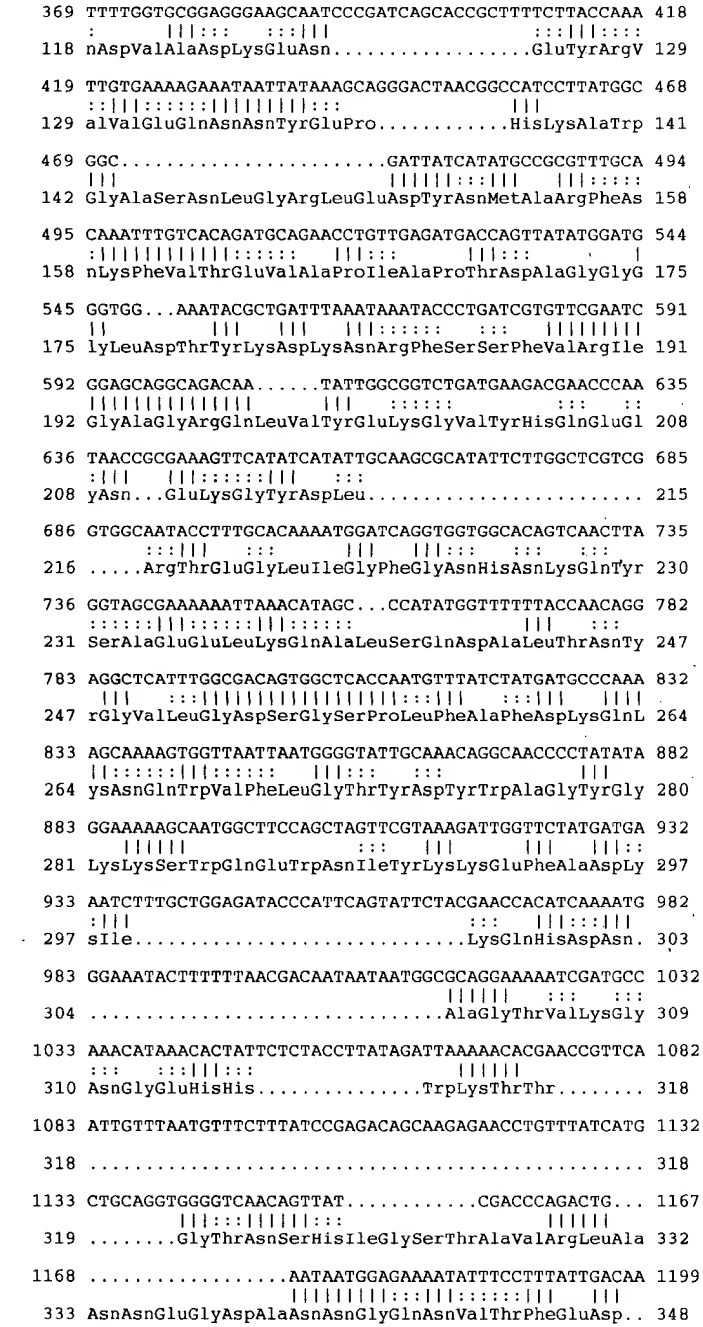
seq_documentation_block:
;Patent No. 5268270
; APPLICANT: Meyer, Thomas F.;Halter, Roman;Pohlner, Johnnnes
; TITLE OF INVENTION: PROCESS FOR PRODUCING PROTEINS USING GRAM
;NEGATIVE HOST CELLS
; NUMBER OF SEQUENCES: 6
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/171,872
; FILING DATE: 01-JUL-1987
;SEQ ID NO:2:
; LENGTH: 1507
5268270-2

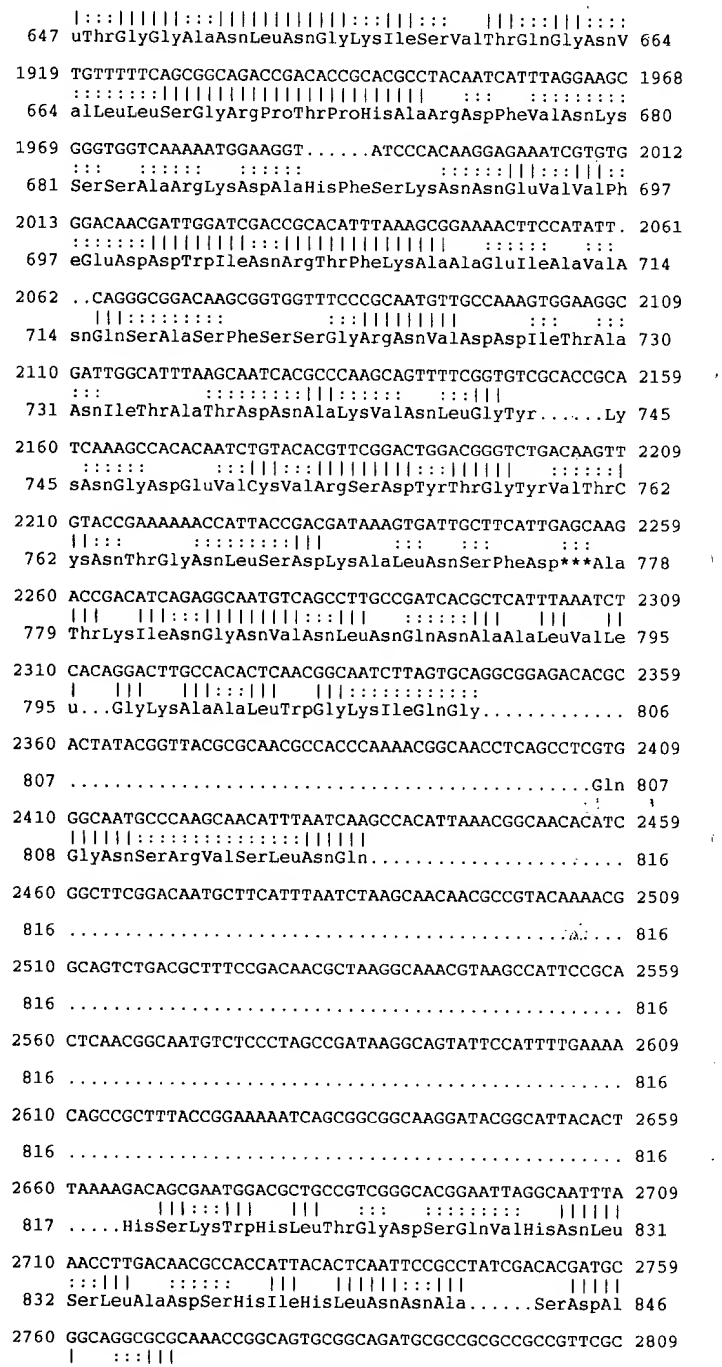
alignment_scores:
Quality: 1515.00 Length: 1725
Ratio: 1.733 Gaps: 44
Percent Similarity: 50.667 Percent Identity: 26.783

alignment_block:
US-09-303-518D-653 x 5268270-2 ..

Align seg 1/1 to: 5268270-2 from: 1 to: 1507

```
 49 AAAACCGGCCGCATCCGCTTCTCGCCCGCTTACTTAGCCATATGCCTGTC 98
    |||:::   |||::::::::::::              |||:::|||   |||::
  2 LysAlaLysArgPheLysIleAsnAlaIleSerLeuSerIlePheLeuAl 18

 99 GTTCGGCATTCTGCCCCAAGCCCGGGCGGGACACACTTATTTCGGCATCA 148
    :::::::::   |||   ::::::|||:::   :::         ::::
 18 aTyrAlaLeuThrProTyrSerGluAlaAlaLeuValArgAspAspValA 35

149 ACTACCAATACTATCGCGACTTTGCCGAAAATAAAGGCAAGTTTGCAGTC 198
    ::||||||   :::|||||||||||||||||||||||||||||   |||
 35 spTyrGlnIlePheArgAspPheAlaGluAsnLysGlyLysPhePheVal 51

199 GGGGCGAAAGATATTGAGGTTTACAACAAAAAAGGGGAGTTGGTCGGCAA 248
    ||||||   |||::::::|||   ||||||:::|||:::   :::|||::
 52 GlyAlaThrAspLeuSerValLysAsnLysArgGlyGlnAsnIleGlyAs 68

249 ATCGATGACGAAAGCCCCGATGATTGATTTTTCTGTGGTATCGCGTAAC. 297
    :::::::::::::::|||||||||||||||||||||||||||::::::   |||
 68 nAlaLeuSerAsnValProMetIleAspPheSerValAlaAspValAsnL 85

298 ..GGCGTGGCGGCATTGGCGGGCGATCAATATATTGTGAGCGTGGCACAT 345
       :::|||:::::::::         ||||||   |||||||||   |||
 85 ysArgIleAlaThrValValAspProGlnTyrAlaValSerValLysHis 101

346 ..........................AACGGCGGCTATAACAATGTTGA 368
                                 |||   ||||||:::   ::
102 AlaLysAlaGluValHisThrPheTyrTyrGlyGlnTyrAsnGlyHisAs 118
```




647 uThrGlyGlyAlaAsnLeuAsnGlyLysIleSerValThrGlnGlyAsnV 664
1919 TGTTTTTCAGCGGCAGACCGACACCGCACGCCTACAATCATTTAGGAAGC 1968
664 alLeuLeuSerGlyArgProThrProHisAlaArgAspPheValAsnLys 680
1969 GGGTGGTCAAAAATGGAAGGT......ATCCCACAAGGAGAAATCGTGTG 2012
681 SerSerAlaArgLysAspAlaHisPheSerLysAsnAsnGluValValPh 697
2013 GGACAACGATTGGATCGACCGCACATTTAAAGCGGAAAACTTCCATATT. 2061
697 eGluAspAspTrpIleAsnArgThrPheLysAlaAlaGluIleAlaValA 714
2062 ..CAGGGCGGACAAGCGGTGGTTTCCCGCAATGTTGCCAAAGTGGAAGGC 2109
714 snGlnSerAlaSerPheSerSerGlyArgAsnValAspAspIleThrAla 730
2110 GATTGGCATTTAAGCAATCACGCCCAAGCAGTTTTCGGTGTCGCACCGCA 2159
731 AsnIleThrAlaThrAspAsnAlaLysValAsnLeuGlyTyr.....Ly 745
2160 TCAAAGCCACACAATCTGTACACGTTCGGACTGGACGGGTCTGACAAGTT 2209
745 sAsnGlyAspGluValCysValArgSerAspTyrThrGlyTyrValThrC 762
2210 GTACCGAAAAAACCATTACCGACGATAAAGTGATTGCTTCATTGAGCAAG 2259
762 ysAsnThrGlyAsnLeuSerAspLysAlaLeuAsnSerPheAsp***Ala 778
2260 ACCGACATCAGAGGCAATGTCAGCCTTGCCGATCACGCTCATTTAAATCT 2309
779 ThrLysIleAsnGlyAsnValAsnLeuAsnGlnAsnAlaAlaLeuValLe 795
2310 CACAGGACTTGCCACACTCAACGGCAATCTTAGTGCAGGCGGAGACACGC 2359
795 u...GlyLysAlaAlaLeuTrpGlyLysIleGlnGly..... 806
2360 ACTATACGGTTACGCGCAACGCCACCCAAAACGGCAACCTCAGCCTCGTG 2409
807 .....Gln 807
2410 GGCAATGCCCAAGCAACATTTAATCAAGCCACATTAAACGGCAACACATC 2459
808 GlyAsnSerArgValSerLeuAsnGln..... 816
2460 GGCTTCGGACAATGCTTCATTTAATCTAAGCAACAACGCCGTACAAAACG 2509
816 816
2510 GCAGTCTGACGCTTTCCGACAACGCTAAGGCAAACGTAAGCCATTCCGCA 2559
816 816
2560 CTCAACGGCAATGTCTCCCTAGCCGATAAGGCAGTATTCCATTTTGAAAA 2609
816 816
2610 CAGCCGCTTTACCGGAAAAATCAGCGGCGGCAAGGATACGGCATTACACT 2659
816 816
2660 TAAAAGACAGCGAATGGACGCTGCCGTCGGGCACGGAATTAGGCAATTTA 2709
817 .....HisSerLysTrpHisLeuThrGlyAspSerGlnValHisAsnLeu 831
2710 AACCTTGACAACGCCACCATTACACTCAATTCCGCCTATCGACACGATGC 2759
832 SerLeuAlaAspSerHisIleHisLeuAsnAsnAla......SerAspAl 846
2760 GGCAGGCGCGCAAACCGGCAGTGCGGCAGATGCGCCGCGCCGCCGTTCGC 2809
846 aGlnSerAla..... 849
2810 GCCGTTCCCTATTATCCGTTACGCCGCCAACTTCGGCAGAATCCCGTTTC 2859
850 .....AsnLysTyr 852
2860 AACACGCTGACGGTAAACGGCAAATTGAACGGTCAGGGAACATTCCGCTT 2909
853 HisThrIleLysIleAsn...HisLeuSerGlyAsnGlyHisPheHisTy 868
2910 TATGTCGGAACTCTTCGGCTACCGCAGCGGCAAATTGAAGCTGGCGGAAA 2959
868 rLeuThrAspLeuAlaLysAsnLeuGlyAspLysValLeuValLysGluA 885
2960 GTTCCGAAGGCACTTACACCTTGGCTGTCAACAATACCGGCAACGAACCC 3009
885 laSerAlaGlyHisTyrGlnLeuHisValGlnAsnLysThrGlyGluPro 901
3010 GTAAGTCTCGAGCAATTGACGGTAGTGGAAGGAAAAGACAACACACCGCT 3059
902 AsnGln.....GluGlyLeuAspLeuPheAspAl 911
3060 GTCC.....GAAAATCTTAATTTCACCCTGCAAAACGAAC 3094
911 aSerSerValGlnAspArgSerArgLeuPheValSerLeuAlaAsnHisT 928
3095 ACGTCGATGCCGGCGCATGGCGTTATCAGCTTATCCGCAAAGACGGCGAG 3144
928 yrValAspLeuGlyAlaLeuArgTyrThrIleLysThrGluAsnGlyIle 944
3145 TTCCGCCTGCATAATCCGGTCAAAGAACAAGAGCTTTCCGACAAACTCGG 3194
945 ThrArgLeuTyrAsnProTyrAlaGly.....AsnGlyArgProVa 958
3195 CAAGGCGGGAGAAACAGAGGCCGCCTTGACGGCAAAACAGGCACAACTTG 3244
958 lLysProAlaProSerProAlaAlaAsnThrAlaSerGlnAlaGlnLysA 975
3245 CCGCC..... 3249
975 laThrGlnThrAspGlyAlaGlnIleAlaLysProGlnAsnIleValVal 991
3250 .....AAACAACAGGC 3260
992 AlaProProSerProGlnAlaAsnGlnAlaGluGluLeuArgGlnGlnAl 1008
3261 GGAAAAAGACAACGCG.....CAAAGCCTTGACGCGCTGATTGCGG 3301
1008 aLysAlaGluGlnValLysArgGlnGlnAlaAlaGluAlaGluLysValA 1025
3302 CCGGGCGCAATGCCACCGAAAAGGCAGAAAGTGTTGCCGAACCGGCCCGG 3351
1025 laArgGlnLysAspGluGluAlaLysArgLysAlaAlaGluIleAlaArg 1041
3352 CAG.....GCAGGCGGGGAAAATGCCGGCATTATGCAGGC 3386
1042 GlnGlnGluGluAlaArgLysAlaAlaGluLeuAlaAlaLysGlnLysAl 1058
3387 GGAGGAAGAGAAAAAACGGGTGCAGGCGGATAAAGACACCGCCTTGGCGA 3436
1058 aGluAlaGluArgLysAlaArgGlu.....LeuAlaA 1069
3437 AACAGCGCGAAGCGGAAACCCGGCCGGCTACCACCGCCTTCCCCCGCGCC 3486
1069 rgGlnLysAlaGluGluAlaSerHisGlnAlaAsnAlaLysProLysArg 1085
3487 CGCCGCGCCCGCCGGGATTTGCCGCAACCG..... 3516
1086 ArgArgArgArgAlaIleLeuProArgProProAlaProValPheSerLe 1102
3516 3516
1102 uAspAspTyrAspAlaLysAspAsnSerGluSerSerIleGlyAsnLeuA 1119

```
3516 ..................................................  3516

1119 laArgValProIleArgMetGlyArgGluLeuIleAsnAspTyrGluGlu 1135

3516 ..................................................  3516

1136 IleProLeuGluGluLeuGluAspGluAlaGluGluGluArgArgGlnAl 1152

3516 ..................................................  3516

1152 aThrGlnPheHisSerLysSerArgAsnArgArgAlaIleSerSerGluP 1169

3516 ..................................................  3516

1169 roSerSerAspGluAspAlaSerGluSerValSerThrSerAspLysHis 1185

3516 ..................................................  3516

1186 ProGlnAspAsnThrGluLeuHisGluLysValGluThrAlaGlyLeuGl 1202

3517 .............CAGCCCCAACCGCAACCCCAACCGCAGCGCGACCTGA 3553
                 ||||||:::   |||         |||   |||   :
1202 nProArgAlaAlaGlnProArgThrGlnAlaAlaAlaGlnAlaAspAlaV 1219

3554 TCAGCCGTTATGCCAATAGCGGTTTGAGTGAATTTTCCGCCACGCTCAAC 3603
     ::|||      :::||||||:::||||||:::      |||:::  :::
1219 alSerThrAsnThrAsnSerAlaLeuSerAspAlaMetAlaSerThrGln 1235

3604 AGCGTT.................TTCGCCGTACAGGACGAATTGGACCG 3635
     |||:::                 :::   ::::::        :::::
1236 SerIleLeuLeuAspThrGlyAlaTyrLeuThrArgHisIleAlaGlnLy 1252

3636 CGTGTTTGCCGAAGACCGCCGCAACGCCGTTTGGACAAGCGGCATCCGGG 3685
     :      |||:::   ::::::|||:::||||||   ::::::
1252 sSerArgAlaAspAlaGluLysAsnSerValTrpMetGluAsnThrGlyT 1269

3686 ACACCAAACACTACCGTTCGCAAGATTTCCGCGCCTACCGCCAACAAACC 3735
          :::   |||   |||   ::::::|||   :::   ::::::
1269 yrGlyArgAspTyrAlaSerAlaGlnTyrArgArgPheSerSerLysArg 1285

3736 GACCTGCGCCAAATCGGTATGCAGAAAAACCTCGGCAGCGGG...CGCGT 3782
               |||||||||:::::::::::|||:::::::::   :::::
1286 ThrGlnThrGlnIleGlyIleAspArgSerLeuSerGluAsnMetGlnIl 1302

3783 CGGCATCCTGTTTTCGCACAACCGGACCGGAAACACCTTCGACGACGGCA 3832
     :|||   ::::::::::::::   :::   :::||||||||||:::::
1302 eGlyGlyValLeuThrTyrSerAspSerGlnHisThrPheAspGlnAla. 1318

3833 TCGGCAACTCGGCACGGCTTGCCCACGGTGCCGTTTTCGGGCAATACGGC 3882
       |||::::::      ::::::::::::   ::::::|||:::|||
1319 ..GlyGlyLysAsnThrPheValGlnAlaAsnLeuTyrGlyLysTyrTyr 1334

3883 ATCGGC...AGGTTCGACATCGGCATCAGCGCGGGCGCGGGTTTTAGTAG 3929
     ::::::     :::   ::::::   :::   |||||||||
1335 LeuAsnAspAlaTrpTyrValAlaGlyAspIleGlyAlaGlySerLeuAr 1351

3930 CGGCAGCCTTTCAGACGGCATCAGAGGCAAAATCCGCCGCCGCCGTGCTGC 3979
      :::   |||:::          ::::::::::::::|||      ::::
1351 gSerArgLeuGlnThrGlnGlnLysAlaAsnPheAsnArgThrSerIleG 1368

3980 ATTACGGCATTCAGGCAAGATACCGCGCAGGTTTCGGCGGATTCGGCATC 4029
     ::   |||:::                    ::::::   |||   |||
1368 lnThrGlyLeuThrLeuGlyAsnThrLeuLysIleAsnGlnPheGluIle 1384

4030 GAACCGCACATCGGCGCAACGCGCTATTTCGTCCAAAAAGCGGATTACCG 4079
       |||       |||              :::::::::||||||||::
1385 ValProSerAlaGlyIleArgTyrSerArgLeuSerSerAlaAspTyrLy 1401

4080 ATACGAAAACGTCAATATCGCCACCCCGGGCCTTGCATTCAACCGCTACC 4129
     :      :::   ::::::   :::   ::::::|||   :::
1401 sLeuGlyAspAspSerValLysValSerSerMetAlaValLysThrLeuT 1418
```

```
4130 GCGCGGGCATTAAGGCAGATTATTCATTCAAACCGGCGCAACACATTTCC 4179
       ||||||:::          |||   ||||||   :::   :::::::::
1418 hrAlaGlyLeuAspPheAlaTyrArgPheLys...ValGlyAsnLeuThr 1433

4180 ATCACGCCTTATTTGAGCCTGTCCTATACCGATGCCGCTTCCGGCAAAGT 4229
     :::   |||   ||||||   :::|||      |||      ||||||
1434 ValLysProLeuLeuSerAlaAlaTyrPhe...AlaAsnTyrGlyLysGl 1449

4230 CCGAACGCGCGTCAATACCGCCCGTATTGGCGCAGGATTTCGGCAAAACCC 4279
          ::::::|||:::         :::|||            :::   :
1449 yGlyValAsnValGlyGlyLysSerPheAlaTyrLysAlaAspAsnGlnG 1466

4280 GCAGTGCGGAATGGGGCGTAAACGCCGAAATCAAAGGTTTCACGCTGTCC 4329
     :::::   :::   ||||||            ::::::   ||||||:::
1466 lnGlnTyrSerAlaGlyValAlaLeuLeuTyrArgAsnValThrLeuAsn 1482

4330 CTCCACGCTGCCGCCGCCAAGGGGCCGCAATTGGAAGCGCAGCACAGCGC 4379
     ::::::::::::   :::||||||   ||||||||   |||   |||::
1483 ValAsnGlySerIleThrLysGlyLysGlnLeuGluLysGlnLysSerGl 1499

4380 GGGCATCAAATTAGGCTACCGCTGG 4404
     :   ||||||:::       |||:::
1499 yGlnIleLysIleGlnIleArgPhe 1507
```

seq_name: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-08-296-791-6

```
seq_documentation_block:
; Sequence 6, Application US/08296791
; Patent No. 6245337
;  GENERAL INFORMATION:
;    APPLICANT:  St. Geme III, Joseph W.
;    APPLICANT:  Falkow, Stanley
;    TITLE OF INVENTION:  Haemophilus Adherence and Penetration
;    TITLE OF INVENTION:  Protein
;    NUMBER OF SEQUENCES:  9
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Flehr, Hohbach, Test, Albritton & Herbert
;      STREET:  4 Embarcadero Center, Suite 3400
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  United States
;      ZIP:  94111-4187
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/296,791
;      FILING DATE:  25-AUG-1994
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Trecartin, Richard F.
;      REGISTRATION NUMBER:  31,801
;      REFERENCE/DOCKET NUMBER:  A-59941/RFT/RMS
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 781-1989
;      TELEFAX:  (415) 398-3249
;      TELEX:  910 277299
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1848 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  unknown
US-08-296-791-6


alignment_scores:
            Quality: 1513.00              Length:   1916
              Ratio:  1.519                Gaps:     60
 Percent Similarity:  51.983   Percent Identity: 25.418

alignment_block:
```

US-09-303-518D-653 x US-08-296-791-6 ..

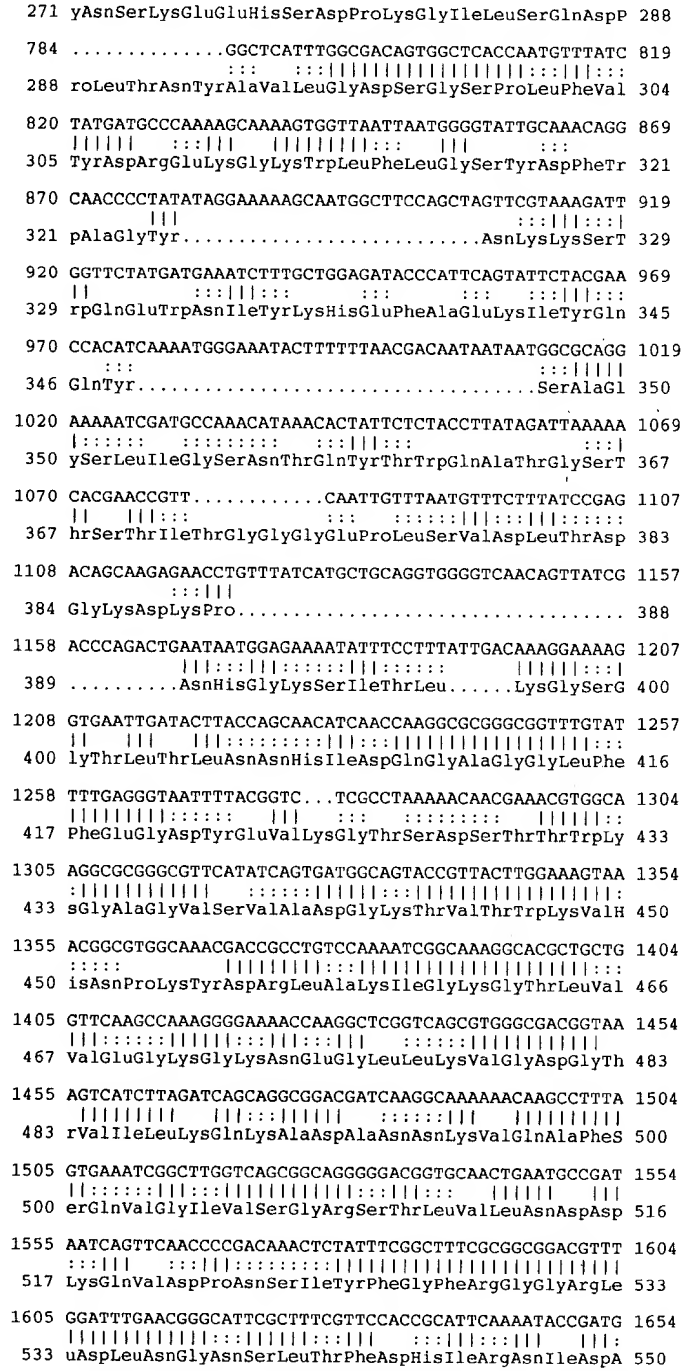
Align seg 1/1 to: US-08-296-791-6 from: 1 to: 1848

```
 64 CGCTTCTCGCCCGCTTACTTAGCCATATGCCTGTCGTTCGGCATTCTGCC 113
    :::|||:::      :::::::|||:::   :::::::::::::::   ||
  5 LysPheLysLeuAsnPheIleAlaLeuThrValAlaTyrAlaLeuThrPr 21

114 CCAAGCCCGGGCGGGACACACTTATTTCGGCATCAACTACCAATACTATC 163
    |   ::::::|||:::   :::          ::::::||||||  ::::|
 21 oTyrThrGluAlaAlaLeuValArgAspAspValAspTyrGlnIlePheA 38

164 GCGACTTTGCCGAAAATAAAGGCAAGTTTGCAGTCGGGGCGAAAGATATT 213
    ||||||||||||||||||||||||||||:::|||||||||   ::::::
 38 rgAspPheAlaGluAsnLysGlyLysPheSerValGlyAlaThrAsnVal 54

214 GAGGTTTACAACAAAAAAGGGGAGTTGGTCGGCAAATCGATGACGAAAGC 263
    ||||||   :::|||||||::::::   :::|||:::::::::   :|:::
 55 GluValArgAspLysLysAsnGlnSerLeuGlySerAlaLeuProAsnGl 71

264 C...CCGATGATTGATTTTTCTGTGGTATCG...CGTAACGGCGTGGCGG 307
    :  ||||||||||||||||||||||||:::      :::   :::|||:
 71 yIleProMetIleAspPheSerValValAspValAspLysArgIleAlaT 88

308 CATTGGCGGGCGATCAATATATTGTGAGCGTGGCACAT...AACGGCGGC 354
    ::|||::::::   ||||||:::|||:::|||   |||   :::::::||
 88 hrLeuValAsnProGlnTyrValValGlyValLysHisValSerAsnGly 104

355 TATAACAATGTTGATTTTGGT.............GCGGAGGGAAGCAA 389
       :::::::::   ||||||                :::::::::||
105 ValSerGluLeuHisPheGlyAsnLeuAsnGlyAsnMetAsnAsnGlyAs 121

390 TCCCGATCAGCACCGC.................TTTTCTTACCAAATTG 421
    |      :::||||||                        |||      |
121 nAlaLysSerHisArgAspValSerSerGluGluAsnArgTyrTyrThrV 138

422 TGAAAAGAAATAATTAT.........................AAAGCA 444
    ||:::::|||||:::                          |||
138 alGluLysAsnAsnPheProThrGluAsnValThrSerPheThrLysGlu 154

445 GGGACTAACGGCCATCCTTATGGCGGCGATTATCATATGCCGCGTTTGCA 494
          :::::::::          ||||||:::||||||||||||
155 GluGlnAspAlaGlnLysArgArgGluAspTyrTyrMetProArgLeuAs 171

495 CAAATTTGTCACAGATGCAGAACCTGTTGAG...ATGACCAGTTATATGG 541
     ||||||||||||:::::   |||:::|||      |||:::      :
171 pLysPheValThrGluValAlaProIleGluAlaSerThrAlaAsnAsnA 188

542 ATGGGTGGAAATACGCTGATTTAAATAAATACCCTGATCGTGTTCGAATC 591
    ::      :::|||   :::   :::|||||||||      ||||||:::
188 snLysGlyGluTyrAsnAsnSerAspLysTyrProAlaPheValArgLeu 204

592 GGAGCAGGCAGACAATAT............................TG 611
    |||:::|||   |||:::
205 GlySerGlyThrGlnPheIleTyrLysLysGlySerArgTyrGlnLeuIl 221

612 GCGGTCTGATGAAGACGAACCCAATAACCGCGAAAGTTCATATCAT.... 657
        :::::::::|||:::   :::|||         ::::::
221 eLeuThrGluLysAspLysGlnGlyAsnLeuLeuArgAsnTrpAspValG 238

658 ....................ATTGCAAGCGCATATTCTTGGCTCGTCGGT 687
                        :::::::::||||||::::::   ::::::
238 lyGlyAspAsnLeuGluLeuValGlyAsnAlaTyrThrTyrGlyIleAla 254

688 GGCAATACCTTTGCACAAAATGGATCAGGTGGTGGCACAGTCAACTTAGG 737
    |||:::   :::      |||   :::::::::|||   :::::::::||
255 GlyThrProTyrLysValAsnHisGluAsnAsnGlyLeuIleGlyPheGl 271

738 TAGCGAAAAAATTAAACATAGC...CCATATGGTTTTTTACCAACAGGA. 783
    |::::::|||   :::|||||   |||   |||:::|||
271 yAsnSerLysGluGluHisSerAspProLysGlyIleLeuSerGlnAspP 288

784 .............GGCTCATTTGGCGACAGTGGCTCACCAATGTTTATC 819
```

```
                  :::   :::|||||||||||||||||||:::|||:::
288 roLeuThrAsnTyrAlaValLeuGlyAspSerGlySerProLeuPheVal 304

820 TATGATGCCCAAAAGCAAAAGTGGTTAATTAATGGGGTATTGCAAACAGG 869
    ||||||   :::|||   |||||||||:::   |||      :::
305 TyrAspArgGluLysGlyLysTrpLeuPheLeuGlySerTyrAspPheTr 321

870 CAACCCCTATATAGGAAAAAGCAATGGCTTCCAGCTAGTTCGTAAAGATT 919
          |||                              :::|||:::|
321 pAlaGlyTyr.........................AsnLysLysSerT 329

920 GGTTCTATGATGAAATCTTTGCTGGAGATACCCATTCAGTATTCTACGAA 969
    ||         :::|||:::      :::      :::   :::|||:::
329 rpGlnGluTrpAsnIleTyrLysHisGluPheAlaGluLysIleTyrGln 345

970 CCACATCAAAATGGGAAATACTTTTTTAACGACAATAATAATGGCGCAGG 1019
    :::                                        :::|||||
346 GlnTyr.................................SerAlaGl 350

1020 AAAAATCGATGCCAAACATAAACACTATTCTCTACCTTATAGATTAAAAA 1069
     |:::::   :::::::::   :::|||:::          :::|
350 ySerLeuIleGlySerAsnThrGlnTyrThrTrpGlnAlaThrGlySerT 367

1070 CACGAACCGTT............CAATTGTTTAATGTTTCTTTATCCGAG 1107
     ||   |||:::           :::   :::::::|||:::|||:::::
367 hrSerThrIleThrGlyGlyGlyGluProLeuSerValAspLeuThrAsp 383

1108 ACAGCAAGAGAACCTGTTTATCATGCTGCAGGTGGGGTCAACAGTTATCG 1157
           :::|||
384 GlyLysAspLysPro..................................... 388

1158 ACCCAGACTGAATAATGGAGAAAATATTTCCTTTATTGACAAAGGAAAAG 1207
           |||:::|||:::::::|||::::::      ||||||:::|
389 ..........AsnHisGlyLysSerIleThrLeu......LysGlySerG 400

1208 GTGAATTGATACTTACCAGCAACATCAACCAAGGCGCGGGCGGTTTGTAT 1257
     ||   |||   |||:::::::::|||:::||||||||||||||||:::
400 lyThrLeuThrLeuAsnAsnHisIleAspGlnGlyAlaGlyGlyLeuPhe 416

1258 TTTGAGGGTAATTTTACGGTC...TCGCCTAAAAACAACGAAACGTGGCA 1304
     |||||||||::::::   |||   :::   :::::::::   ||||||::
417 PheGluGlyAspTyrGluValLysGlyThrSerAspSerThrThrTrpLy 433

1305 AGGCGCGGGCGTTCATATCAGTGATGGCAGTACCGTTACTTGGAAAGTAA 1354
     :|||||||||||||   ::::::||||||:::||||||||||||||||:
433 sGlyAlaGlyValSerValAlaAspGlyLysThrValThrTrpLysValH 450

1355 ACGGCGTGGCAAACGACCGCCTGTCCAAAATCGGCAAAGGCACGCTGCTG 1404
     :::::          |||||||||:::||||||||||||||||||||:::
450 isAsnProLysTyrAspArgLeuAlaLysIleGlyLysGlyThrLeuVal 466

1405 GTTCAAGCCAAAGGGGAAAACCAAGGCTCGGTCAGCGTGGGCGACGGTAA 1454
     |||::::::|||||||:::|||:::|||   :::::|||||||||||||
467 ValGluGlyLysGlyLysAsnGluGlyLeuLeuLysValGlyAspGlyTh 483

1455 AGTCATCTTAGATCAGCAGGCGGACGATCAAGGCAAAAAACAAGCCTTTA 1504
      |||||||||   |||:::||||||   :::::::|||   ||||||||||
483 rValIleLeuLysGlnLysAlaAspAlaAsnAsnLysValGlnAlaPheS 500

1505 GTGAAATCGGCTTGGTCAGCGGCAGGGGGACGGTGCAACTGAATGCCGAT 1554
     ||::::::|||:::||||||||||||:::|||:::   ||||||   |||
500 erGlnValGlyIleValSerGlyArgSerThrLeuValLeuAsnAspAsp 516

1555 AATCAGTTCAACCCCGACAAACTCTATTTCGGCTTTCGCGGCGGACGTTT 1604
     :::|||   :::|||:::::::::|||||||||||||||||||||||||
517 LysGlnValAspProAsnSerIleTyrPheGlyPheArgGlyGlyArgLe 533

1605 GGATTTGAACGGGCATTCGCTTTCGTTCCACCGCATTCAAAATACCGATG 1654
     ||||||||||||:::||||||:::|||   :::|||:::|||   |||:
533 uAspLeuAsnGlyAsnSerLeuThrPheAspHisIleArgAsnIleAspA 550

1655 AAGGGGCGATGATTGTCAACCACAATCAAGACAAAGAATCCACCGTTACC 1704
     ::|||||   :::|||||||||||||:::   :::   |||::::::|||
```

```
    1735 TyrLeuSerAsnAlaAspPheAlaLeuAlaGlnAspArgIleLysValAs 1751

    4107 GGGCCTTGCATTCAACCGCTACCGCGCGGGCATTAAGGCAGATTATTCAT 4156
         ::::::   :::         |||   :::      :::|||:::::
    1751 nProIleSerValLysThrAlaPheAlaGlnValAspLeuSerTyrThrT 1768

    4157 TCAAACCGGCGCAACACATTTCCATCACGCCTTATTTGAGCCTGTCCTAT 4206
         ::      ::::::   :::|||||||||||||   ||||||      |||
    1768 yrHisLeuGlyGlu...PheSerIleThrProIleLeuSerAlaArgTyr 1783

    4207 ACCGATGCCGCTTCCGGCAAAGTCCGAACGCGCGTCAATACCGCCGTATT 4256
           ||||||   :::|||:::   :::   :::|||::::::      ::
    1784 ...AspAlaAsnGlnGlyAsnGlyLysIleAsnValSerValTyrAspPh 1799

    4257 GGCGCAGGATTTCGGCAAAACCCGCAGTGCGGAATGGGGCGTAAACGCCG 4306
         :|||   :::      :::   ::::::   :::   |||::::::   :
    1799 eAlaTyrAsnValGluAsnGlnGlnGlnTyrAsnAlaGlyLeuLysLeuL 1816

    4307 AAATCAAAGGTTTCACGCTGTCCCTCCACGCTGCCGCCGCCAAGGGGCCG 4356
         ::      :::      |||||||||   ::::::   :::|||:::
    1816 ysTyrHisAsnValLysLeuSerLeuIleGlyGlyLeuThrLysAlaLys 1832

    4357 CAATTGGAAGCGCAGCACAGCGCGGGCATCAAATTAGGCTAC 4398
         |||   |||   |||   :::|||   :::||||||::::::
    1833 GlnAlaGluLysGlnLysThrAlaGluValLysLeuSerPhe 1846

seq_name: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:PCT-US95-10661A-6

seq_documentation_block:
; Sequence 6, Application PC/TUS9510661A
;  GENERAL INFORMATION:
;    APPLICANT:  Washington University, et al.
;    TITLE OF INVENTION:  Haemophilus Adherence and Penetration Protein
;    NUMBER OF SEQUENCES:  9
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Flehr, Hohbach, Test, Albritton & Herbert
;      STREET:  4 Embarcadero Center, Suite 3400
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  United States
;      ZIP:  94111-4187
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US95/10661A
;      FILING DATE:  16-AUG-1995
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/296,791
;      FILING DATE:  25-AUG-1994
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Trecartin, Richard F.
;      REGISTRATION NUMBER:  31,801
;      REFERENCE/DOCKET NUMBER:  FP-59941/RFT
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 781-1989
;      TELEFAX:  (415) 398-3249
;      TELEX:  910 277299
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1848 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  unknown
PCT-US95-10661A-6


alignment_scores:
           Quality: 1513.00            Length:   1916
             Ratio:  1.519               Gaps:     60
```

```
 Percent Similarity:  51.983   Percent Identity: 25.418

alignment_block:
 US-09-303-518D-653 x PCT-US95-10661A-6   ..

 Align seg 1/1  to: PCT-US95-10661A-6  from: 1  to: 1848

      64 CGCTTCTCGCCCGCTTACTTAGCCATATGCCTGTCGTTCGGCATTCTGCC 113
         :::|||:::      :::::::|||:::   ::::::::::::::   ||
       5 LysPheLysLeuAsnPheIleAlaLeuThrValAlaTyrAlaLeuThrPr 21

     114 CCAAGCCCGGGCGGGACACACTTATTTCGGCATCAACTACCAATACTATC 163
         |   ::::::|||:::   :::         ::::::||||||   :::|
      21 oTyrThrGluAlaAlaLeuValArgAspAspValAspTyrGlnIlePheA 38

     164 GCGACTTTGCCGAAAATAAAGGCAAGTTTGCAGTCGGGGCGAAAGATATT 213
         ||||||||||||||||||||||||||||||:::|||||||||   ::::::
      38 rgAspPheAlaGluAsnLysGlyLysPheSerValGlyAlaThrAsnVal 54

     214 GAGGTTTACAACAAAAAAGGGGAGTTGGTCGGCAAATCGATGACGAAAGC 263
         ||||||   :::||||||::::::   :::|||::::::::::   :::::
      55 GluValArgAspLysLysAsnGlnSerLeuGlySerAlaLeuProAsnGl 71

     264 C...CCGATGATTGATTTTTCTGTGGTATCG...CGTAACGGCGTGGCGG 307
         :   |||||||||||||||||||||||:::      :::   :::|||:
      71 yIleProMetIleAspPheSerValValAspValAspLysArgIleAlaT 88

     308 CATTGGCGGGCGATCAATATATTGTGAGCGTGGCACAT...AACGGCGGC 354
         ::|||::::::   ||||||:::|||:::|||   |||   ::::::|||
      88 hrLeuValAsnProGlnTyrValValGlyValLysHisValSerAsnGly 104

     355 TATAACAATGTTGATTTTGGT.............GCGGAGGGAAGCAA 389
            :::::::::   ||||||                  :::::::::||
     105 ValSerGluLeuHisPheGlyAsnLeuAsnGlyAsnMetAsnAsnGlyAs 121

     390 TCCCGATCAGCACCGC.................TTTCTTACCAAATTG 421
         |      :::||||||                      |||      |
     121 nAlaLysSerHisArgAspValSerSerGluGluAsnArgTyrTyrThrV 138

     422 TGAAAAGAAATAATTAT.........................AAAGCA 444
         ||:::::|||||:::                          |||
     138 alGluLysAsnAsnPheProThrGluAsnValThrSerPheThrLysGlu 154

     445 GGGACTAACGGCCATCCTTATGGCGGCGATTATCATATGCCGCGTTTGCA 494
               :::::::::            ||||||:::||||||||||||
     155 GluGlnAspAlaGlnLysArgArgGluAspTyrTyrMetProArgLeuAs 171

     495 CAAATTTGTCACAGATGCAGAACCTGTTGAG...ATGACCAGTTATATGG 541
          ||||||||||||:::::   |||:::|||      |||:::      :
     171 pLysPheValThrGluValAlaProIleGluAlaSerThrAlaAsnAsnA 188

     542 ATGGGTGGAAATACGCTGATTTAAATAAATACCCTGATCGTGTTCGAATC 591
         ::      :::|||   :::   :::|||||||||      ||||||:::
     188 snLysGlyGluTyrAsnAsnSerAspLysTyrProAlaPheValArgLeu 204

     592 GGAGCAGGCAGACAATAT.............:.............TG 611
         |||:::|||   |||:::
     205 GlySerGlyThrGlnPheIleTyrLysLysGlySerArgTyrGlnLeuIl 221

     612 GCGGTCTGATGAAGACGAACCCAATAACCGCGAAAGTTCATATCAT.... 657
             :::::::::|||:::   :::|||         :::::
     221 eLeuThrGluLysAspLysGlnGlyAsnLeuLeuArgAsnTrpAspValG 238

     658 ...................ATTGCAAGCGCATATTCTTGGCTCGTCGGT 687
                             :::::::::||||||::::::   ::::::
     238 lyGlyAspAsnLeuGluLeuValGlyAsnAlaTyrThrTyrGlyIleAla 254

     688 GGCAATACCTTTGCACAAAATGGATCAGGTGGTGGCACAGTCAACTTAGG 737
         |||:::   :::      |||   :::::::::|||   :::::::::||
     255 GlyThrProTyrLysValAsnHisGluAsnAsnGlyLeuIleGlyPheGl 271

     738 TAGCGAAAAAATTAAACATAGC...CCATATGGTTTTTTACCAACAGGA. 783
         |:::::::|||   :::||||||   |||   |||:::|||
```


271 yAsnSerLysGluGluHisSerAspProLysGlyIleLeuSerGlnAspP 288
784 .....GGCTCATTTGGCGACAGTGGCTCACCAATGTTTATC 819
288 roLeuThrAsnTyrAlaValLeuGlySerGlySerProLeuPheVal 304
820 TATGATGCCCAAAAGCAAAAGTGGTTAATTAATGGGGTATTGCAAACAGG 869
305 TyrAspArgGluLysGlyLysTrpLeuPheLeuGlySerTyrAspPheTr 321
870 CAACCCCTATATAGGAAAAAGCAATGGCTTCCAGCTAGTTCGTAAAGATT 919
321 pAlaGlyTyr.....AsnLysLysSerT 329
920 GGTTCTATGATGAAATCTTTGCTGGAGATACCCATTCAGTATTCTACGAA 969
329 rpGlnGluTrpAsnIleTyrLysHisGluPheAlaGluLysIleTyrGln 345
970 CCACATCAAAATGGGAAATACTTTTTTAACGACAATAATAATGGCGCAGG 1019
346 GlnTyr.....SerAlaGl 350
1020 AAAAATCGATGCCAAACATAAACACTATTCTCTACCTTATAGATTAAAAA 1069
350 ySerLeuIleGlySerAsnThrGlnTyrThrTrpGlnAlaThrGlySerT 367
1070 CACGAACCGTT.....CAATTGTTTAATGTTTCTTTATCCGAG 1107
367 hrSerThrIleThrGlyGlyGluProLeuSerValAspLeuThrAsp 383
1108 ACAGCAAGAGAACCTGTTTATCATGCTGCAGGTGGGGTCAACAGTTATCG 1157
384 GlyLysAspLysPro..... 388
1158 ACCCAGACTGAATAATGGAGAAAATATTTCCTTTATTGACAAAGGAAAAG 1207
389 .....AsnHisGlyLysSerIleThrLeu.....LysGlySerG 400
1208 GTGAATTGATACTTACCAGCAACATCAACCAAGGCGCGGGCGGTTTGTAT 1257
400 lyThrLeuThrLeuAsnAsnHisIleAspGlnGlyAlaGlyGlyLeuPhe 416
1258 TTTGAGGGTAATTTTACGGTC...TCGCCTAAAAACAACGAAACGTGGCA 1304
417 PheGluGlyAspTyrGluValLysGlyThrSerAspSerThrThrTrpLy 433
1305 AGGCGCGGGCGTTCATATCAGTGATGGCAGTACCGTTACTTGGAAAGTAA 1354
433 sGlyAlaGlyValSerValAlaAspGlyLysThrValThrTrpLysValH 450
1355 ACGGCGTGGCAAACGACCGCCTGTCCAAAATCGGCAAAGGCACGCTGCTG 1404
450 isAsnProLysTyrAspArgLeuAlaLysIleGlyLysGlyThrLeuVal 466
1405 GTTCAAGCCAAAGGGGAAAACCAAGGCTCGGTCAGCGTGGGCGACGGTAA 1454
467 ValGluGlyLysGlyLysAsnGluGlyLeuLeuLysValGlyAspGlyTh 483
1455 AGTCATCTTAGATCAGCAGGCGGACGATCAAGGCAAAAAACAAGCCTTTA 1504
483 rValIleLeuLysGlnLysAlaAspAlaAsnAsnLysValGlnAlaPheS 500
1505 GTGAAATCGGCTTGGTCAGCGGCAGGGGGACGGTGCAACTGAATGCCGAT 1554
500 erGlnValGlyIleValSerGlyArgSerThrLeuValLeuAsnAspAsp 516
1555 AATCAGTTCAACCCCGACAAACTCTATTTCGGCTTTCGCGGCGGACGTTT 1604
517 LysGlnValAspProAsnSerIleTyrPheGlyPheArgGlyGlyArgLe 533
1605 GGATTTGAACGGGCATTCGCTTTCGTTCCACCGCATTCAAAATACCGATG 1654
533 uAspLeuAsnGlyAsnSerLeuThrPheAspHisIleArgAsnIleAspA 550
1655 AAGGGGCGATGATTGTCAACCACAATCAAGACAAAGAATCCACCGTTACC 1704
550 spGlyAlaArgValValAsnHisAsnMetThrAsnThrSerAsnIleThr 566
1705 ATTACAGGCAATAAAGATATTACTACA.....ACCGGCAATAA 1742
567 IleThrGlyGluSerLeuIleThrAsnProAsnThrIleThrSerTyrAs 583
1743 C..... 1743
583 nIleGluAlaGlnAspAspAspHisProLeuArgIleArgSerIleProT 600
1743 1743
600 yrArgGlnLeuTyrPheAsnGlnAspAsnArgSerTyrTyrThrLeuLys 616
1743 1743
617 LysGlyAlaSerThrArgSerGluLeuProGlnAsnSerGlyGluSerAs 633
1743 1743
633 nGluAsnTrpLeuTyrMetGlyArgThrSerAspAlaAlaLysArgAsnV 650
1744 .....AACAACTTGGATAGCAAAAAAGAAATTGCCTACAACGGTTGGTTT 1788
650 alMetAsnHisIleAsnAsnGluArgMetAsnGlyPheAsnGlyTyrPhe 666
1789 GGCGAGAAAGATGCAACCAAAACG...AACGGGCGGCTCAATCTGAATTA 1835
667 GlyGluGluGluThrLysAlaThrGlnAsnGlyLysLeuAsnValThrPh 683
1836 CCAACCGGAAGAAGCGGATCGCACTTTACTGCTTTCCGGCGGAACAAATT 1885
683 eAsnGlyLysSerAspGlnAsnArgPheLeuLeuThrGlyGlyThrAsnL 700
1886 TAAACGGCAATATCACGCAAACAAACGGCAAACTGTTTTTCAGCGGCAGA 1935
700 euAsnGlyAspLeuAsnValGluLysGlyThrLeuPheLeuSerGlyArg 716
1936 CCGACACCGCACGCCTACAATCATTTAGGAAGCGGGTGGTCAAAAATGGA 1985
717 ProThrProHisAlaArgAspIleAlaGlyIleSerSerThrLysLysAs 733
1986 AGGT.....ATCCCACAAGGAGAAATCGTGTGGGACAACGATTGGATCG 2029
733 pProHisPheThrGluAsnAsnGluValValValGluAspAspTrpIleA 750
2030 ACCGCACATTTAAAGCGGAAAACTTCCATATTCAGGGCGGACAAGCGGTG 2079
750 snArgAsnPheLysAlaThrThrMetAsnValThrGlyAsnAlaSerLeu 766
2080 GTTTCC...CGCAATGTTGCCAAAGTGGAAGGCGATTGGCATTTAAGCAA 2126
767 TyrSerGlyArgAsnValAlaAsnIleThrSerAsnIleThrAlaSerAs 783
2127 TCACGCCCAAGCAGTTTTCGGTGTCGCACCGCATCAAAGCCACACAATCT 2176
783 nAsnAlaGlnValHisIleGlyTyr.....LysThrGlyAspThrValC 798
2177 GTACACGTTCGGACTGGACGGGTCTGACAAGTTGTACCGAAAAAACCATT 2226
798 ysValArgSerAspTyrThrGlyTyrValThrCysHisAsnSerAsnLeu 814
2227 ACCGACGATAAAGTGATTGCTTCATTGAGCAAGACCGACATCAGAGGCAA 2276
815 Ser...GluLysAlaLeuAsnSerPheAsnProThrAsnLeuArgGlyAs 830
2277 TGTCAGCCTTGCCGATCACGCTCATTTAAATCTCACAGGACTTGCCACAC 2326
830 nValAsnLeuThrGluAsnAlaSerPheThrLeu...GlyLysAlaAsnL 846


2327 TCAACGGCAATCTTAGTGCAGGCGGA 2352
846 euPheGlyThrIleGlnSerIleGlyThrSerGlnValAsnLeuLysGlu 862
2353 GACACGCACTATACGGTTACGCGCAACGCCACC 2385
863 AsnSerHisTrpHisLeuThrGlyAsnSerAsnValAsnGlnLeuAsnLe 879
2386 .CAAAACGGCAACCTCAGCCTCGTGGGCAATGCCCAA 2421
879 uThrAsnGlyHisIleHisLeu.....AsnAlaGlnAsnAspAlaAsnL 894
2422 .....GCAACATTTAATCAAGCCACATTAAACGGCAACACATCGGCTTCG 2466
894 ysValThrThrTyrAsnThrLeuThrValAsn.....SerLeuSer 907
2467 GACAATGCTTCATTT.....AATCTAAGCAA 2492
908 GlyAsnGlySerPheTyrTyrTrpValAspPheThrAsnAsnLysSerAs 924
2493 CAACGCCGTACAAAACGGCAGT 2514
924 nLysValValValAsnLysSerAlaThrGlyAsnPheThrLeuGlnValA 941
2515 .....CTGACGCTTTCCGACAAC 2532
941 laAspLysThrGlyGluProAsnHisAsnGluLeuThrLeuPheAspAla 957
2533 GCTAAGGCA.....AACGTAAGCCATTCCGCACTCAACGGCAATGT 2573
958 SerAsnAlaThrArgAsnAsnLeuGluValThrLeuAlaAsnGlySerVa 974
2574 CTCCCTAGCCGATAAGGCAGTATTCCATTTTGAAAACAGCCGCTTTACCG 2623
974 l.....AspArgGlyAlaTrpLysTyrLysLeuArgAsnValAsnG 988
2624 GAAAAATCAGCGGCGGCAAGGATACGGCATTACACTTAAAAGACAGCGAA 2673
988 lyArgTyrAsp.....LeuTyrAsnProGlu 996
2674 TGGACGCTGCCGTCGGGCACGGAATTAGGCAATTTAAACCTTGACAACGC 2723
997 .....ValGluLysArgAsnGlnThrValAspThrTh 1007
2724 CACCATTACACTCAATTCCGCCTATCGACACGATGCGGCAGGCGCGCAAA 2773
1007 rAsnIleThrThrProAsnAspIleGlnAlaAspAlaProSerAlaGlnS 1024
2774 CCGGCAGT 2781
1024 erAsnAsnGluGluIleAlaArgValGluThrProValProProProAla 1040
2782 .....GCGGCAGATGCGCCGCGCCGCCGT..... 2805
1041 ProAlaThrGluSerAlaIleAlaSerGluGlnProGluThrArgProAl 1057
2805 2805
1057 aGluThrAlaGlnProAlaMetGluGluThrAsnThrAlaAsnSerThrG 1074
2806 .....TCGCGC 2811
1074 luThrAlaProLysSerAspThrAlaThrGlnThrGluAsnProAsnSer 1090
2812 CGTTCCCTATTATCCGTTACG 2832
1091 GluSerValProSerGluThrThrGluLysValAlaGluAsnProProGl 1107
2832 2832
1107 nGluAsnGluThrValAlaLysAsnGluGlnGluAlaThrGluProThrP 1124
2833 .....CCGCCAACTTCGGCAGAA 2850
1124 roGlnAsnGlyGluValAlaLysGluAspGlnProThrValGluAlaAsn 1140
2851 TCCCGTTTCAACACGCTGACGGTAAAC...GGCAAATTGAACGGTCAGGG 2897
1141 ThrGlnThrAsnGluAlaThrGlnSerGluGlyLysThrGluGluThrGl 1157
2898 AACATTCCGCTTTATGTCGGAA 2919
1157 nThrAlaGluThrLysSerGluProThrGluSerValThrValSerGluA 1174
2920 .....CTCTTCGGC 2928
1174 snGlnProGluLysThrValSerGlnSerThrGluAspLysValValVal 1190
2929 TACCGCAGCGGCAAATTGAAGCTG 2952
1191 GluLysGluGluLysAlaLysValGluThrGluGluThrGlnLysAlaPr 1207
2952 2952
1207 oGlnValThrSerLysGluProProLysGlnAlaGluProAlaProGluG 1224
2953 .....GCGGAAAGTTCCGAAGGCACT 2973
1224 luValProThrAspThrAsnAlaGluGluAlaGlnAlaLeuGlnGlnThr 1240
2974 TACACCTTGGCTGTC 2988
1241 GlnProThrThrValAlaAlaAlaGluThrThrSerProAsnSerLysPr 1257
2988 2988
1257 oAlaGluGluThrGlnGlnProSerGluLysThrAsnAlaGluProValT 1274
2989 .....AACAATACCGGCAACGAACCCGTAAGTCTCGAGCAA 3024
1274 hrProValValSerGluAsnThrAlaThrGlnProThrGluThrGluGlu 1290
3025 TTGACGGTAGTGGAAGGAAAAGACAACACACCGCTGTCCGAAAATCTTAA 3074
1291 ThrAlaLysValGlu.LysGluLysThrGlnGluValProGlnVal.... 1305
3075 TTTCACCCTGCAAAACGAACACGTCGATGCCGGCGCATGGCGTTATCAGC 3124
1306 .....AlaSerGlnGluSerPro.LysGlnGluGlnProAlaAl 1318
3125 TTATCCGCAAAGACGGCGAGTTCCGCCTGCATAATCCGGTCAAAGAACAA 3174
1318 aLysProGlnAlaGlnThrLysProGlnAla.GluProAlaArgGluAsn 1334
3175 GAGCTTTCCGACAAA 3189
1335 ValLeuThrThrLysAsnValGlyGluProGlnProGlnAlaGlnProGl 1351
3190 .....CTCGGCAAGGCGGGAGAAACAGAGGCC...GCCT 3220
1351 nThrGlnSerThrAlaValProThrThrGlyGluThrAlaAlaAsnSerL 1368
3221 TGACGGCAAAACAGGCACAACTTGCCGCCAAACAACAGGCGGAA..... 3264
1368 ysProAlaAlaLysProGlnAlaGlnAlaLysProGlnThrGluProAla 1384
3265 .....AAAGACAACGCGCAAAGCCTTGA 3287
1385 ArgGluAsnValSerThrValAsnThrLysGluProGlnSerGlnThrSe 1401
3288 CGCGCTGATTGCGGCCGGGCGCAATGCCACCGAAAAGGCAGAAAGTGTTG 3337
1401 rAlaThrValSerThrGluGlnProAlaLysGluThrSerSerAsnValG 1418
3338 CCGAACCGGCC 3348

```
1418 luGlnProAlaProGluAsnSerIleAsnThrGlySerAlaThrThrMet 1434

3348 ..................................................  3348

1435 ThrGluThrAlaGluLysSerAspLysProGlnMetGluThrValThrGl 1451

3349 .......CGGCAG.................GCAGGCGGGGAAAATGCCG 3373
             ||||||                 |||   ::::::   |||:
1451 uAsnAspArgGlnProGluAlaAsnThrValAlaAspAsnSerValAlaA 1468

3374 GCATTATGCAGGCGGAGGAAGAGAAAAAACGG.....GTGCAG 3411
     ::      ::::::::::|||:::|||:::|||       |||::
1468 snAsnSerGluSerSerGluSerLysSerArgArgArgArgSerValSer 1484

3412 GCGGATAAAGACACCGCCTTGGCGAAACAGCGCGAAGCGGAAACCCGGCC 3461
          |||:::|||:::      :::         |||:::|||:::
1485 GlnProLysGluThrSerAlaGluGluThrThrValAlaSerThrGlnGl 1501

3462 GGCTACC.................ACCGCCTTCCCCCGCGCCCGCCGCG 3493
      :::|||                 |||      ||||||:::||||||:
1501 uThrThrValAspAsnSerValSerThrProLysProArgSerArgArgT 1518

3494 CCCGCCGGGATTTGCCGCAACCGCAGCCCCAACCGCAACCCCAACCGCAG 3543
     ::||||||:::::         :::   :::|||          |||
1518 hrArgArgSerValGlnThrAsnSerTyrGluProValGluLeuProThr 1534

3543 ..................................................  3543

1535 GluAsnAlaGluAsnAlaGluAsnValGlnSerGlyAsnAsnValAlaAs 1551

3544 ..............CGCGACCTGATCAGCCGTTATGCCAATAGCGGTT 3577
                  |||:::|||   |||:::   :::|||:::   :
1551 nSerGlnProAlaLeuArgAsnLeuThrSerLysAsnThrAsnAlaValI 1568

3578 TGAGTGAATTTTCCGCC..............ACGCTCAACAGCGTTTTC 3612
     ::|||:::      |||               :::||||||
1568 leSerAsnAlaMetAlaLysAlaGlnPheValAlaLeuAsnValGlyLys 1584

3613 GCCGTACAGGACGAATTGGACCGCGTG...TTTGCCGAAGACCGCCGCAA 3659
     ||||||:::::::::::::::::::::   :::   ::::::   :::
1585 AlaValSerGlnHisIleSerGlnLeuGluMetAsnAsnGluGlyGlnTy 1601

3660 CGCCGTTTGGACAAGCGGCATCCGGGACACCAAACACTACCGTTCGCAAG 3709
          ||||||   |||:::          :::||||:::|||   |||::::
1601 rAsnValTrpIleSerAsnThrSerMetAsnLysAsnTyrSerSerGluG 1618

3710 ATTTCCGCGCCTACCGCCAACAAACCGACCTGCGCCAAATCGGTATGCAG 3759
     :::::|||   :::   :::::::::         |||:::|||   :::
1618 lnTyrArgArgPheSerSerLysSerThrGlnThrGlnLeuGlyTrpAsp 1634

3760 AAAAACCTCGGCAGCGGG...CGCGTCGGCATCCTGTTTTCGCACAACCG 3806
     ::::::::::::::::::   :::::|||   :::|||::::::   ||
1635 GlnThrIleSerAsnAsnValGlnLeuGlyGlyValPheThrTyrValAr 1651

3807 GACCGGAAACACCTTCGACGACGGCATCGGCAACTCGGCACGGCTTGCCC 3856
     |:::::|||:::||||||   :::   ::::::::::::
1651 gAsnSerAsnAsnPheAspLysAlaSerSerLysAsnThrLeuAlaGlnV 1668

3857 ACGGTGCCGTTTTCGGGCAATACGGCATCGGCAGGTTCGACATCGGCATC 3906
       :::                |||:::   :::::::::   :::||||||
1668 alAsnPheTyrSerLysTyrTyrAlaAspAsnHisTrpTyrLeuGlyIle 1684

3907 AGCGCGGGCGCGGGTTTTAGTAGCGGCAGCCTTTCAGACGGCATCAGAGG 3956
     :::   |||   |||      ::::::::::|||:::   :::   :::::
1685 AspLeuGlyTyrGlyLysPheGlnSerAsnLeuGlnThrAsnAsnAsnAl 1701

3957 CAAAATCCGCCGCCGCGTGCTGCATTACGGCATTCAGGCAAGATACCGCG 4006
     :|||:::   |||::::::   :::   |||:::   |||
1701 aLysPheAlaArgHisThrAlaGlnIleGlyLeuThrAlaGlyLysAlaP 1718

4007 CAGGTTTCGGCGGATTCGGCATCGAACCGCACATCGGCGCAACGCGCTAT 4056
       ::::::|||:::|||::::::::::|||   :::|||:::
1718 heAsnLeuGlyAsnPheAlaValLysProThrValGlyValArgTyrSer 1734

4057 TTCGTCCAAAAAGCGGATTACCGATACGAAAACGTCAATATCGCCACCCC 4106
     ::::::::::::|||||||:::          :::   :::|||   :::
1735 TyrLeuSerAsnAlaAspPheAlaLeuAlaGlnAspArgIleLysValAs 1751

4107 GGGCCTTGCATTCAACCGCTACCGCGCGGGCATTAAGGCAGATTATTCAT 4156
          ::::::   :::          |||   :::      :::|||::::
1751 nProIleSerValLysThrAlaPheAlaGlnValAspLeuSerTyrThrT 1768

4157 TCAAACCGGCGCAACACATTTCCATCACGCCTTATTTGAGCCTGTCCTAT 4206
     ::      ::::::   :::||||||||||||   ||||||      |||
1768 yrHisLeuGlyGlu...PheSerIleThrProIleLeuSerAlaArgTyr 1783

4207 ACCGATGCCGCTTCCGGCAAAGTCCGAACGCGCGTCAATACCGCCGTATT 4256
        ||||||   :::|||:::   :::   :::|||::::::      ::
1784 ...AspAlaAsnGlnGlyAsnGlyLysIleAsnValSerValTyrAspPh 1799

4257 GGCGCAGGATTTCGGCAAAACCCGCAGTGCGGAATGGGGCGTAAACGCCG 4306
     :|||   :::      :::   ::::::   :::   |||::::::   :
1799 eAlaTyrAsnValGluAsnGlnGlnGlnTyrAsnAlaGlyLeuLysLeuL 1816

4307 AAATCAAAGGTTTCACGCTGTCCCTCCACGCTGCCGCCGCCAAGGGGCCG 4356
     ::      :::      |||||||||   ::::::   :::|||:::
1816 ysTyrHisAsnValLysLeuSerLeuIleGlyGlyLeuThrLysAlaLys 1832

4357 CAATTGGAAGCGCAGCACAGCGCGGGCATCAAATTAGGCTAC 4398
     |||   |||   |||   :::|||   :::||||||::::::
1833 GlnAlaGluLysGlnLysThrAlaGluValLysLeuSerPhe 1846
```

seq_name: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:US-08-728-470-9

```
seq_documentation_block:
; Sequence 9, Application US/08728470
; Patent No. 5928651
;  GENERAL INFORMATION:
;    APPLICANT:  Barenkamp, Stephen J
;    TITLE OF INVENTION:  High Molecular Weight Surface Proteins
;    TITLE OF INVENTION:  of No. 5928651-Typeable Haemophilus
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Shoemaker and Mattare, Ltd.
;      STREET:  2001 Jefferson Davis Hwy., 1203 Crystal Plaza
;      STREET:  Bldg. 1
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22202-0286
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/728,470
;      FILING DATE:
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/302,832
;      FILING DATE:  16-MAR-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US PCT/US93/02166
;      FILING DATE:  16-MAR-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9205704.1
;      FILING DATE:  16-MAR-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Berkstresser, Jerry W
;      REGISTRATION NUMBER:  22,651
;      REFERENCE/DOCKET NUMBER:  1038-633
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 415-0810
;      TELEFAX:  (703) 415-0813
;  INFORMATION FOR SEQ ID NO:  9:
```

```
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1338 amino acids
;       TYPE:  amino acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
US-08-728-470-9


alignment_scores:
            Quality:  308.50              Length:   1542
              Ratio:  0.439                Gaps:     75
 Percent Similarity:  45.525   Percent Identity: 20.169

alignment_block:
 US-09-303-518D-653 x US-08-728-470-9   ..

 Align seg 1/1  to: US-08-728-470-9  from: 1  to: 1338

       1 ATGAAAACAACCGACAAACGGACAACCGAAACACACCGCAAAGCCCCTAA 50
         |||   |||   ||||||   |||   :::|||
      67 MetIleThrGlyAspLysValThrLeuLysThrGlyAlaValIleAspLe 83

      51 AACCGGCCGCATCCGCTTCTCGCCCGCTTACTTAGCCATATGCCTGTCGT 100
          :::|||:::
      83 uSerGlyLys..................................... 86

     101 TCGGCATTCTGCCCCAAGCCCGGGCGGGACACACTTATTTCGGCATCAAC 150
                          :::::::|||:::|||||||:::|||
      87 ...................GluGlyGlyGluThrTyrLeuGly..... 94

     151 TACCAATACTATCGCGACTTTGCCGAAAATAAAGGCAAGTTTGCAGTCGG 200
                              :::::::::::|||:::         ||
      95 ....................GlyAspGluArgGlyGlu.........Gl 101

     201 GGCGAAAGATATTGAGGTTTACAACAAAAAAGGGGAGTTGGTCGGCAAAT 250
         |   :::   |||::::::   :::|||                    :
     101 yLysAsnGlyIleGlnLeuAlaLysLys.................ThrT 112

     251 CGATGACGAAAGCCCCGATGATTGATTTTTCTGTGGTATCGCGTAACGGC 300
         :::::   |||:::      |||:::   |||      ::::::::::|||
     112 hrLeuGluLysGlySerThrIleAsnValSerGlyLysGluLysGlyGly 128

     301 GTGGCGGCATTGGCGGGCGATCAATATATTGTG................. 333
         |||   :::   ||||||      ::::::
     129 ArgAlaIleValTrpGlyAspIleAlaLeuIleAspGlyAsnIleAsnAl 145

     334 ..........AGCGTGGCACATAACGGCGGC................... 354
                   ::::::|||   :::||||||
     145 aGlnGlyLysAspIleAlaLysThrGlyGlyPheValGluThrSerGlyH 162

     355 ..TATAACAATGTTGATTTTGGTGCGGAGGGAAGCAATCCC.......... 393
           |||   ::::::|||   :::|||      ::::::
     162 isTyrLeuSerIleAspAspAsnAlaIleValLysThrLysGluTrpLeu 178

     394 ...GATCAGCACCGCTTTTCTTACCAAATTGTGAAAAGAAATAATTATAA 440
           |||   ::::::   :::   :::      :::   ::::::   ::
     179 LeuAspProGluAsnValThrIleGluAlaProSerAlaSerArgValGl 195

     441 AGCAGGGACTAACGGCCATCCTTATGGCGGCGATTATCATATGCCGCGTT 490
         :   |||::::::   :::   ::::::::::::              |
     195 uLeuGlyAlaAspArgAsnSerHisSerAlaGluValIleLysValThrL 212

     491 TGCACAAATTTGTCACAGATGCAGAACCTGTTGAGATGACCAGTTATATG 540
         ||   |||      |||:::         :::      |||:::
     212 euLysLysAsnAsnThrSerLeuThrThrLeuThrAsnThrThrIleSer 228

     541 GATGGGTGGAAATACGCTGATTTAAATAAATACCCTGATCGTGTTCGAAT 590
         :::      |||   |||   :::   :::         |||   :::::
     229 AsnLeuLeuLysSerAlaHisValValAsnIleThrAlaArgArgLysLe 245

     591 C.............................................. 591
         :
     245 uThrValAsnSerSerIleSerIleGluArgGlySerHisLeuIleLeuH 262

     592 ........GGAGCAGGCAGACAATATTGGCGGTCTGATGAAGACGAACCC 633
                 |||   |||   |||      :::   |||:::|||
     262 isSerGluGlyGlnGlyGlyGlnGlyValGlnIleAspLysAsp...... 276

     634 AATAACCGCGAAAGTTCATATCATATTGCAAGCGCATATTCTTGGCTCGT 683
                                  |||:::|||
     277 ........................IleThrSerGlu.............. 280

     684 CGGTGGCAATACCTTTGCACAAAATGGATCAGGTGGT.............. 720
          |||||||||                |||||||||
     281 .GlyGlyAsnLeuThrIleTyr......SerGlyGlyTrpValAspValH 295

     721 ..GGCACAGTCAACTTAGGTAGCGAAAAAATTAAACATAGCCCATATGGT 768
             :::::::::|||||||||                          |||
     295 isLysAsnIleThrLeuGlySer.......................Gly 303

     769 TTTTTACCAACAGGAGGCTCATTTGGCGAC.................AG 800
         ||||||            :::   ||||||                ||
     304 PheLeuAsnIleThrThrLysGluGlyAspIleAlaPheGluAspLysSe 320

     801 TGGCTCACCAATGTTTATCTATGATGCCCAAAAGCAAAAGTGGTTAATTA 850
         ||||         :::         ||||||
     320 rGlyArgAsnAsnLeuThrIleThrAlaGln................... 330

     851 ATGGGGTATTGCAAACAGGCAACCCCTATATAGGAAAAAGCAATGGCTTC 900
           |||:::::::   :::||||||                |||||||||||||
     331 ..GlyThrIleThrSerGlyAsn..............SerAsnGlyPhe 341

     901 CAGCTAGTTCGTAAAGATTGGTTCTATGATGAAATCTTTGCTGGAGATAC 950
         :::                          ::::::::::::      ::::::
     342 Arg.....................PheAsnAsnValSerLeuAsnSerLe 351

     951 CCATTCAGTATTCTACGAACCACATCAAAATGGGAAATACTTTTTTAACG 1000
                                     :::||||||      |||:::|
     351 u..........................GlyGlyLysLeuSerPheThrA 359

    1001 ACAATAATAATGGCGCAGGAAAA........................... 1023
         ||:::::::::      |||:::
     359 spSerArgGluAspArgGlyArgArgThrLysGlyAsnIleSerAsnLys 375

    1024 ................................ATCGATGCCAAACA 1037
                                         |||:::   |||
     376 PheAspGlyThrLeuAsnIleSerGlyThrValAspIleSerMetLysAl 392

    1038 TAAACACTATTCTCTACCTTATAGATTAAAAACACGAACCGTT....... 1080
                |||       ||||||   |||   ||||||
     392 aProLysValSerTrpPheTyrArgAspLysGlyArgThrTyrTrpAsnV 409

    1081 .........................CAATTGTTTAATGTTTCTTTATCC 1104
                                    :::   ||||||:::|||::::::
     409 alThrThrLeuAsnValThrSerGlySerLysPheAsnLeuSerIleAsp 425

    1105 GAGACAGCAAGAGAACCTGTTTATCATGCTGCAGGTGGGGTCAACAGTTA 1154
         :::|||                              |||   :::|||
     426 SerThr............................GlySerGlySerTh 432

    1155 TCGACCCAGACTGAATAATGGAGAAAATATTTCCTTTATTGACAAAGGAA 1204
             |||   ::::::|||:::|||   ::::::   |||      :::|
     432 rGlyProSerIleArgAsnAlaGlu...LeuAsnGlyIleThrPheAsnL 448

    1205 AAGGTGAATTGATACTTACCAGCAACATCAACCAAGGCGCGGGCGGTTTG 1254
         ||:::            |||   ||||||   ||||||:::
     448 ysAla............ThrPheAsnIleAlaGlnGlySer......... 457

    1255 TATTTTGAGGGTAATTTTACGGTCTCGCCT..............AAAAA 1289
                ::::||||||:::::::::                  ||||::
     458 ......ThrAlaAsnPheSerIleLysAlaSerIleMetProPheLysSe 472

    1290 CAACGAAACGTGGCAA......GGCGCGGGCGTTCATATCAGTGATGGCA 1333
         :|||   ::::::         :::      :::   :::|||   |||:
     472 rAsnAlaAsnTyrAlaLeuPheAsnGluAspIleSerValSerGlyGlyG 489
```

```
1005 ........AsnIleThrThrLysThrGlyAspIleAsnGlyLys...... 1016

2902 TTCCGCTTTATGTCGGAACTCTTCGGCTACCGCAGCGGCAAATTGAAGCT 2951

1016 ..................................................... 1016

2952 GGCGGAAAGTTCCGAAGGCACTTAC.....................ACCT 2980
      :::||||||||:::|||:::                          ||||
1017 .ValGluSerSerSerGlySerValThrLeuValAlaThrGlyAlaThrL 1033

2981 TGGCTGTCAACAATACCGGCAACGAACCCGTAAGTCTCGAGCAATTGACG 3030
     |||||||:::|||   :::::::::   |||:::          :::|||
1033 euAlaValGlyAsnIleSerGlyAsnThrValThr.........IleThr 1046

3031 GTAGTGGAAGGAAAAGACAACACACCGCTGTCCGAAAATCTTAATTTCAC 3080
     :::   :::||||||   ::::::   ::::::::::::::::|||   ||
1047 AlaAspSerGlyLysLeuThrSerThrValGlySerThrIleAsnGlyTh 1063

3081 CCTGCAAAACGAACACGTCGATGCCGGCGCATGGCGTTATCAGCTTATCC 3130
     |
1063 r..................................................... 1063

3131 GCAAAGACGGCGAGTTCCGCCTGCATAATCCGGTCAAAGAACAAGAGCTT 3180
                          ::::::   :::   :::::::::
1064 .....................AsnSerValThrThrSerSerGlnSer 1072

3181 TCCGACAAACTCGGCAAGGCGGGAGAAACAGAGGCCGCCTTGACGGCAAA 3230
     :::|||     |||     :::  :::   :::   :::||||||::
1073 GlyAspIleGluGlyThrIleSerGlyAsnThrValAsnValThrAlaSe 1089

3231 ACAGGCACAACTT............GCCGCCAAACAACAGGCGGAAAAAG 3268
     :   :::::|||           :::||||||   :::|||   |||:
1089 rThrGlyAspLeuThrIleGlyAsnSerAlaLysValGluAla...LysA 1105

3269 ACAACGCGCAAAGCCTTGACGCGCTGATTGCGGCCGGGCGCAATGCCACC 3318
     :::::|||   :::|||   |||         :::|||:::   :::|||
1105 snGlyAlaAlaThrLeuThrAla......GluSerGlyLysLeuThrThr 1119

3319 GAAAAGGCAGAAAGTGTT............................GC 3338
     :::   :::::::|||:::                            ||
1120 GlnThrGlySerSerIleThrSerSerAsnGlyGlnThrThrLeuThrAl 1136

3339 CGAACCGGCCCGGCAGGCAGGCGGGGAAAATGCC............... 3372
     |:::   :::      ||||||:::   ||||||
1136 aLysAspSerSerIleAlaGlyAsnIleAsnAlaAlaAsnValThrLeuA 1153

3373 ........GGCATTATGCAGGCGGAGGAAGAGAAAAAACGGGTGCAGGCG 3414
             |||   :::   :::      :::::::|||   ::::::|||
1153 snThrThrGlyThrLeuThrThrThrGlyAspSerLys...IleAsnAla 1168

3415 GATAAAGACACCGCCTTGGCGAAACAGCGCGAAGCGGAAACCCGGCCGGC 3464
        :::   |||       :::   ::::::|||:::          ||
1169 ThrSerGlyThrLeuThrIleAsnAlaLysAspAlaLysLeuAspGlyAl 1185

3465 TACCACCGCCTTCCCCCGCGCCCGCCGCGCCCGCCGGGATTTGCCGCAAC 3514
     |:::::::::
1185 aAlaSerGly..................................... 1188

3515 CGCAGCCCCAACCGCAACCCCAACCGCAGCGCGACCTGATCAGCCGTTAT 3564
                               :::|||   :::::::::
1189 .........................AspArgThrValValAsnAlaThr 1196

3565 GCCAATAGCGGTTTGAGTGAATTTTCCGCCACGCTCAACAGCGTTTTCGC 3614
           ||||||   ::::::   :::|||      :::|||
1197 AsnAlaSerGlySerGlyAsnValThrAlaLysThrSerSerSerValAs 1213

3615 CGTACAGGACGAATTGGAC..............CGCGTGTTTGCCGAAG 3649
      :::      :::|||:::              :::::::::::::|||:
1213 nIleThrGlyAspLeuAsnThrIleAsnGlyLeuAsnIleIleSerGluA 1230

3650 ACCGCCGCAACGCCGTTTGGACAAGCGGCATCCGG...GACACCAAACAC 3696
     ::   ||||||:::|||         |||   :::   |||:::|||:::
1230 snGlyArgAsnThrValArgLeuArgGlyLysGluIleAspValLysTyr 1246

3697 TACCGTTCGCAAGATTTCCGCGCCTACCGCCAACAAACCGACCTGCGCCA 3746
               |||    ::::::   ::::::::::::          ||||
1247 Ile......GlnPro.GlyValAlaSerValGluGluValIleGluAlaL 1261

3747 AATCGGTATGCAGAAAAACC.........................TCG 3769
     ||   |||   :::|||:::                          :::
1261 ysArgValLeuGluLysValLysAspLeuSerAspGluGluArgGluThr 1277

3770 GCAGCGGGCGCGTCGGCATCCTGTTTTCGCACAACCGGACCGGAAACACC 3819
        |||       ::::::|||:::   |||   :::   |||::::::
1278 LeuAlaLysLeuGlyValSerAlaValArgPheValGluProAsnAsnAl 1294

3820 TTCGACGACGGCATCGGCAACTCGGCACGGCTTGCCCACGGTGCCGTTTT 3869
        |||:::         |||:::::::   :::   |||   |||   |
1294 aIleThrVal......AsnThrGlnAsnGluPheThrThrLysProSerS 1309

3870 CGGGCAATACGGCATCGGCAGGTTCGACA......TCGGCATCAGCGCGG 3913
     ||      |||   |||   |||:::::::      |||::::::
1309 erGlnValThrIleSerGluGlyLysAlaCysPheSerSerGlyAsnGly 1325

3914 GCGCGGGTTTTAGTAGCGGCAGCC 3937
     |||||||||   :::   :::|||
1326 AlaArgValCysThrAsnValAla 1333
```

seq_name: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-08-719-641-9

seq_documentation_block:

; Sequence 9, Application US/08719641
; Patent No. 6218141
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen J
; TITLE OF INVENTION: High Molecular Weight Surface Proteins
; TITLE OF INVENTION: of No. 6218141-Typeable Haemophilus
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Mattare, Ltd.
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; STREET: Bldg. 1
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/719,641
; FILING DATE:
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/302,832
; FILING DATE: 16-SEP-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US PCT/US93/02166
; FILING DATE: 16-MAR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9205704.1
; FILING DATE: 16-MAR-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Berkstresser, Jerry W
; REGISTRATION NUMBER: 22,651
; REFERENCE/DOCKET NUMBER: 1038-625
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 415-0813
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1338 amino acids
; TYPE: amino acid
; STRANDEDNESS: single

```
;      TOPOLOGY:  linear
US-08-719-641-9


alignment_scores:
           Quality:  308.50              Length:   1542
             Ratio:  0.439                Gaps:     75
 Percent Similarity:  45.525   Percent Identity: 20.169

alignment_block:
 US-09-303-518D-653 x US-08-719-641-9   ..

 Align seg 1/1  to: US-08-719-641-9  from: 1  to: 1338

       1 ATGAAAACAACCGACAAACGGACAACCGAAACACACCGCAAAGCCCCTAA 50
         |||   |||   ||||||   |||   :::|||
      67 MetIleThrGlyAspLysValThrLeuLysThrGlyAlaValIleAspLe 83

      51 AACCGGCCGCATCCGCTTCTCGCCCGCTTACTTAGCCATATGCCTGTCGT 100
          :::|||:::
      83 uSerGlyLys..................................... 86

     101 TCGGCATTCTGCCCCAAGCCCGGGCGGGACACACTTATTTCGGCATCAAC 150
                        ::::::|||:::||||||:::|||
      87 ...................GluGlyGlyGluThrTyrLeuGly...... 94

     151 TACCAATACTATCGCGACTTTGCCGAAAATAAAGGCAAGTTTGCAGTCGG 200
                          ::::::::::::|||:::         ||
      95 ....................GlyAspGluArgGlyGlu.........Gl 101

     201 GGCGAAAGATATTGAGGTTTACAACAAAAAAGGGGAGTTGGTCGGCAAAT 250
         |   :::   |||::::::   :::|||                  :
     101 yLysAsnGlyIleGlnLeuAlaLysLys.................ThrT 112

     251 CGATGACGAAAGCCCCGATGATTGATTTTTCTGTGGTATCGCGTAACGGC 300
         :::::   |||:::      |||:::   |||   ' ::::::::|||
     112 hrLeuGluLysGlySerThrIleAsnValSerGlyLysGluLysGlyGly 128

     301 GTGGCGGCATTGGCGGGCGATCAATATATTGTG................. 333
            |||   :::   ||||||      ::::::
     129 ArgAlaIleValTrpGlyAspIleAlaLeuIleAspGlyAsnIleAsnAl 145

     334 ..........AGCGTGGCACATAACGGCGGC.................... 354
                  :::::::|||   :::||||||
     145 aGlnGlyLysAspIleAlaLysThrGlyGlyPheValGluThrSerGlyH 162

     355 ..TATAACAATGTTGATTTTGGTGCGGAGGGAAGCAATCCC......... 393
           |||   :::::::|||   :::|||      ::::::
     162 isTyrLeuSerIleAspAspAsnAlaIleValLysThrLysGluTrpLeu 178

     394 ...GATCAGCACCGCTTTTCTTACCAAATTGTGAAAAGAAATAATTATAA 440
            |||   ::::::   :::   :::      :::   ::::::   ::
     179 LeuAspProGluAsnValThrIleGluAlaProSerAlaSerArgValGl 195

     441 AGCAGGGACTAACGGCCATCCTTATGGCGGCGATTATCATATGCCGCGTT 490
         :   |||::::::   :::   ::::::::::::                |
     195 uLeuGlyAlaAspArgAsnSerHisSerAlaGluValIleLysValThrL 212

     491 TGCACAAATTTGTCACAGATGCAGAACCTGTTGAGATGACCAGTTATATG 540
         ||   |||      |||:::         :::      |||:::
     212 euLysLysAsnAsnThrSerLeuThrThrLeuThrAsnThrThrIleSer 228

     541 GATGGGTGGAAATACGCTGATTTAAATAAATACCCTGATCGTGTTCGAAT 590
         :::      |||   |||   :::   :::         |||   :::::
     229 AsnLeuLeuLysSerAlaHisValValAsnIleThrAlaArgArgLysLe 245

     591 C.............................................. 591
         :
     245 uThrValAsnSerSerIleSerIleGluArgGlySerHisLeuIleLeuH 262

     592 ........GGAGCAGGCAGACAATATTGGCGGTCTGATGAAGACGAACCC 633
                 |||   |||   |||      :::   |||:::|||
     262 isSerGluGlyGlnGlyGlyGlnGlyValGlnIleAspLysAsp...... 276

     634 AATAACCGCGAAAGTTCATATCATATTGCAAGCGCATATTCTTGGCTCGT 683
                                 |||:::|||
     277 ........................IleThrSerGlu.............. 280

     684 CGGTGGCAATACCTTTGCACAAAATGGATCAGGTGGT............. 720
          |||||||||                  |||||||||
     281 .GlyGlyAsnLeuThrIleTyr......SerGlyGlyTrpValAspValH 295

     721 ..GGCACAGTCAACTTAGGTAGCGAAAAAATTAAACATAGCCCATATGGT 768
               :::::::::|||||||||                      |||
     295 isLysAsnIleThrLeuGlySer.......................Gly 303

     769 TTTTTACCAACAGGAGGCTCATTTGGCGAC..................AG 800
         ||||||            :::   ||||||                  ||
     304 PheLeuAsnIleThrThrLysGluGlyAspIleAlaPheGluAspLysSe 320

     801 TGGCTCACCAATGTTTATCTATGATGCCCAAAAGCAAAAGTGGTTAATTA 850
         ||||         :::         ||||||
     320 rGlyArgAsnAsnLeuThrIleThrAlaGln................... 330

     851 ATGGGGTATTGCAAACAGGCAACCCCTATATAGGAAAAAGCAATGGCTTC 900
           |||::::::   :::||||||               ||||||||||||
     331 ..GlyThrIleThrSerGlyAsn..............SerAsnGlyPhe 341

     901 CAGCTAGTTCGTAAAGATTGGTTCTATGATGAAATCTTTGCTGGAGATAC 950
         :::                     ::::::::::::      ::::::
     342 Arg.....................PheAsnAsnValSerLeuAsnSerLe 351

     951 CCATTCAGTATTCTACGAACCACATCAAAATGGGAAATACTTTTTTAACG 1000
                                    :::||||||      |||:::|
     351 u.........................GlyGlyLysLeuSerPheThrA 359

    1001 ACAATAATAATGGCGCAGGAAAA.......................... 1023
         ||:::::::::      |||:::
     359 spSerArgGluAspArgGlyArgArgThrLysGlyAsnIleSerAsnLys 375

    1024 ..................................ATCGATGCCAAACA 1037
                                           |||:::   |||
     376 PheAspGlyThrLeuAsnIleSerGlyThrValAspIleSerMetLysAl 392

    1038 TAAACACTATTCTCTACCTTATAGATTAAAAACACGAACCGTT....... 1080
                   |||      ||||||   |||   ||||||
     392 aProLysValSerTrpPheTyrArgAspLysGlyArgThrTyrTrpAsnV 409

    1081 .........................CAATTGTTTAATGTTTCTTTATCC 1104
                                    :::   ||||||:::|||:::::
     409 alThrThrLeuAsnValThrSerGlySerLysPheAsnLeuSerIleAsp 425

    1105 GAGACAGCAAGAGAACCTGTTTATCATGCTGCAGGTGGGGTCAACAGTTA 1154
         :::|||                                |||   :::|||
     426 SerThr............................GlySerGlySerTh 432

    1155 TCGACCCAGACTGAATAATGGAGAAAATATTTCCTTTATTGACAAAGGAA 1204
            |||   ::::::|||:::|||   ::::::   |||      :::|
     432 rGlyProSerIleArgAsnAlaGlu...LeuAsnGlyIleThrPheAsnL 448

    1205 AAGGTGAATTGATACTTACCAGCAACATCAACCAAGGCGCGGGCGGTTTG 1254
         ||:::            |||   ||||||   ||||||:::
     448 ysAla............ThrPheAsnIleAlaGlnGlySer......... 457

    1255 TATTTTGAGGGTAATTTTACGGTCTCGCCT..............AAAAA 1289
                :::||||||:::::::::                    |||::
     458 ......ThrAlaAsnPheSerIleLysAlaSerIleMetProPheLysSe 472

    1290 CAACGAAACGTGGCAA......GGCGCGGGCGTTCATATCAGTGATGGCA 1333
         :|||   ::::::         :::      :::   ::::||   |||:
     472 rAsnAlaAsnTyrAlaLeuPheAsnGluAspIleSerValSerGlyGlyG 489

    1334 GTACCGTTACTTGGAAAGTAAACGGC....................... 1359
         :::::|||:::::::|||:::|||:::
     489 lySerValAsnPheLysLeuAsnAlaSerSerSerAsnIleGlnThrPro 505
```

```
1016 ..................................................  1016

2952 GGCGGAAAGTTCCGAAGGCACTTAC.....................ACCT 2980
      :::||||||||||:::|||:::                        ||||
1017 .ValGluSerSerSerGlySerValThrLeuValAlaThrGlyAlaThrL 1033

2981 TGGCTGTCAACAATACCGGCAACGAACCCGTAAGTCTCGAGCAATTGACG 3030
     ||||||||:::|||   :::::::::   |||:::         :::|||
1033 euAlaValGlyAsnIleSerGlyAsnThrValThr.........IleThr 1046

3031 GTAGTGGAAGGAAAAGACAACACACCGCTGTCCGAAAATCTTAATTTCAC 3080
     :::   :::||||||   ::::::   ::::::::::::::::|||   ||
1047 AlaAspSerGlyLysLeuThrSerThrValGlySerThrIleAsnGlyTh 1063

3081 CCTGCAAAACGAACACGTCGATGCCGGCGCATGGCGTTATCAGCTTATCC 3130
     |
1063 r..................................................  1063

3131 GCAAAGACGGCGAGTTCCGCCTGCATAATCCGGTCAAAGAACAAGAGCTT 3180
                          ::::::   :::   ::::::::
1064 ......................AsnSerValThrThrSerSerGlnSer 1072

3181 TCCGACAAACTCGGCAAGGCGGGAGAAACAGAGGCCGCCTTGACGGCAAA 3230
     :::|||      |||      :::   :::   :::   :::||||||::
1073 GlyAspIleGluGlyThrIleSerGlyAsnThrValAsnValThrAlaSe 1089

3231 ACAGGCACAACTT............GCCGCCAAACAACAGGCGGAAAAAG 3268
     :   :::::::|||           :::||||||   :::|||   |||:
1089 rThrGlyAspLeuThrIleGlyAsnSerAlaLysValGluAla...LysA 1105

3269 ACAACGCGCAAAGCCTTGACGCGCTGATTGCGGCCGGGCGCAATGCCACC 3318
     :::::|||   :::|||   |||         :::|||:::   :::|||
1105 snGlyAlaAlaThrLeuThrAla......GluSerGlyLysLeuThrThr 1119

3319 GAAAAGGCAGAAAGTGTT............................GC 3338
     :::   :::::::|||:::                            ||
1120 GlnThrGlySerSerIleThrSerSerAsnGlyGlnThrThrLeuThrAl 1136

3339 CGAACCGGCCCGGCAGGCAGGCGGGGAAAATGCC.............. 3372
     |:::   :::      ||||||:::   ||||||
1136 aLysAspSerSerIleAlaGlyAsnIleAsnAlaAlaAsnValThrLeuA 1153

3373 ........GGCATTATGCAGGCGGAGGAAGAGAAAAAACGGGTGCAGGCG 3414
             |||   :::   :::       :::::::|||   ::::::|||
1153 snThrThrGlyThrLeuThrThrThrGlyAspSerLys...IleAsnAla 1168

3415 GATAAAGACACCGCCTTGGCGAAACAGCGCGAAGCGGAAACCCGGCCGGC 3464
        :::   |||         :::   ::::::|||:::         ||
1169 ThrSerGlyThrLeuThrIleAsnAlaLysAspAlaLysLeuAspGlyAl 1185

3465 TACCACCGCCTTCCCCCGCGCCCGCCGCGCCCGCCGGGATTTGCCGCAAC 3514
     |:::::::::
1185 aAlaSerGly..................................... 1188

3515 CGCAGCCCCAACCGCAACCCCAACCGCAGCGCGACCTGATCAGCCGTTAT 3564
                              :::|||   :::::::::
1189 .......................AspArgThrValValAsnAlaThr 1196

3565 GCCAATAGCGGTTTGAGTGAATTTTCCGCCACGCTCAACAGCGTTTTCGC 3614
          ||||||   ::::::   :::|||      :::|||
1197 AsnAlaSerGlySerGlyAsnValThrAlaLysThrSerSerSerValAs 1213

3615 CGTACAGGACGAATTGGAC..............CGCGTGTTTGCCGAAG 3649
      :::      :::|||:::               ::::::::::|||:
1213 nIleThrGlyAspLeuAsnThrIleAsnGlyLeuAsnIleIleSerGluA 1230

3650 ACCGCCGCAACGCCGTTTGGACAAGCGGCATCCGG...GACACCAAACAC 3696
     ::   ||||||:::|||         |||   :::   |||:::|||:::
1230 snGlyArgAsnThrValArgLeuArgGlyLysGluIleAspValLysTyr 1246

3697 TACCGTTCGCAAGATTTCCGCGCCTACCGCCAACAAACCGACCTGCGCCA 3746
           |||    ::::::   ::::::::::::       ||||
1247 Ile......GlnPro.GlyValAlaSerValGluGluValIleGluAlaL 1261

     3747 AATCGGTATGCAGAAAAACC.........................TCG 3769
          ||   |||   :::|||:::                        :::
     1261 ysArgValLeuGluLysValLysAspLeuSerAspGluGluArgGluThr 1277

     3770 GCAGCGGGCGCGTCGGCATCCTGTTTTCGCACAACCGGACCGGAAACACC 3819
             |||      ::::::|||:::   |||   :::   |||:::::::
     1278 LeuAlaLysLeuGlyValSerAlaValArgPheValGluProAsnAsnAl 1294

     3820 TTCGACGACGGCATCGGCAACTCGGCACGGCTTGCCCACGGTGCCGTTTT 3869
             |||:::         |||::::::   :::   |||   |||   |
     1294 aIleThrVal......AsnThrGlnAsnGluPheThrThrLysProSerS 1309

     3870 CGGGCAATACGGCATCGGCAGGTTCGACA......TCGGCATCAGCGCGG 3913
          ||      |||   |||   |||::::::      |||::::::
     1309 erGlnValThrIleSerGluGlyLysAlaCysPheSerSerGlyAsnGly 1325

     3914 GCGCGGGTTTTAGTAGCGGCAGCC 3937
          |||||||||   :::   :::|||
     1326 AlaArgValCysThrAsnValAla 1333
```

```
seq_name: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:US-08-617-697-9

seq_documentation_block:
; Sequence 9, Application US/08617697
; Patent No. 5977336
;  GENERAL INFORMATION:
;    APPLICANT:  Barenkamp, Stephen J
;    TITLE OF INVENTION:  High Molecular Weight Surface Proteins
;    TITLE OF INVENTION:  of No. 5977336-Typeable Haemophilus
;    NUMBER OF SEQUENCES:  11
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Shoemaker and Mattare, Ltd.
;      STREET:  2001 Jefferson Davis Hwy., 1203 Crystal Plaza
;      STREET:  Bldg. 1
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22202-0286
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/617,697
;      FILING DATE:  01-APR-1996
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/302,832
;      FILING DATE:  05-OCT-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US PCT/US93/02166
;      FILING DATE:  16-MAR-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Berkstresser, Jerry W
;      REGISTRATION NUMBER:  22,651
;      REFERENCE/DOCKET NUMBER:  1038-557
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 415-0810
;      TELEFAX:  (703) 415-0813
;  INFORMATION FOR SEQ ID NO:  9:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1599 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-617-697-9


alignment_scores:
            Quality:  308.50              Length:   1542
              Ratio:  0.439                Gaps:     75
```

Percent Similarity: 45.525 Percent Identity: 20.169

alignment_block:
US-09-303-518D-653 x US-08-617-697-9 ..

Align seg 1/1 to: US-08-617-697-9 from: 1 to: 1599

```
      1 ATGAAAACAACCGACAAACGGACAACCGAAACACACCGCAAAGCCCCTAA 50
        |||   |||   ||||||   |||   :::|||
    327 MetIleThrGlyAspLysValThrLeuLysThrGlyAlaValIleAspLe 343

     51 AACCGGCCGCATCCGCTTCTCGCCCGCTTACTTAGCCATATGCCTGTCGT 100
         :::|||:::
    343 uSerGlyLys..................................... 346

    101 TCGGCATTCTGCCCCAAGCCCGGGCGGGACACACTTATTTCGGCATCAAC 150
                           :::::::|||:::|||||||:::|||
    347 ....................GluGlyGlyGluThrTyrLeuGly..... 354

    151 TACCAATACTATCGCGACTTTGCCGAAAATAAAGGCAAGTTTGCAGTCGG 200
                             :::::::::::|||:::         ||
    355 .....................GlyAspGluArgGlyGlu.........Gl 361

    201 GGCGAAAGATATTGAGGTTTACAACAAAAAAGGGGAGTTGGTCGGCAAAT 250
        |   :::   |||::::::   :::|||                   :
    361 yLysAsnGlyIleGlnLeuAlaLysLys.................ThrT 372

    251 CGATGACGAAAGCCCCGATGATTGATTTTTCTGTGGTATCGCGTAACGGC 300
        :::::   |||:::      |||:::   |||      ::::::::::|||
    372 hrLeuGluLysGlySerThrIleAsnValSerGlyLysGluLysGlyGly 388

    301 GTGGCGGCATTGGCGGGCGATCAATATATTGTG................. 333
           |||   :::   ||||||      :::::
    389 ArgAlaIleValTrpGlyAspIleAlaLeuIleAspGlyAsnIleAsnAl 405

    334 ..........AGCGTGGCACATAACGGCGGC................... 354
                 :::::::|||   :::||||||
    405 aGlnGlyLysAspIleAlaLysThrGlyGlyPheValGluThrSerGlyH 422

    355 ..TATAACAATGTTGATTTTGGTGCGGAGGGAAGCAATCCC......... 393
          |||   ::::::|||   :::|||      ::::::
    422 isTyrLeuSerIleAspAspAsnAlaIleValLysThrLysGluTrpLeu 438

    394 ...GATCAGCACCGCTTTTCTTACCAAATTGTGAAAAGAAATAATTATAA 440
           |||   ::::::   :::   :::      :::   :::::: . ::
    439 LeuAspProGluAsnValThrIleGluAlaProSerAlaSerArgValGl 455

    441 AGCAGGGACTAACGGCCATCCTTATGGCGGCGATTATCATATGCCGCGTT 490
        :   |||::::::   :::   ::::::::::::                |
    455 uLeuGlyAlaAspArgAsnSerHisSerAlaGluValIleLysValThrL 472

    491 TGCACAAATTTGTCACAGATGCAGAACCTGTTGAGATGACCAGTTATATG 540
        ||   |||      |||:::      :::      |||:::
    472 euLysLysAsnAsnThrSerLeuThrThrLeuThrAsnThrThrIleSer 488

    541 GATGGGTGGAAATACGCTGATTTAAATAAATACCCTGATCGTGTTCGAAT 590
        :::      |||   |||   :::   :::         |||   :::::
    489 AsnLeuLeuLysSerAlaHisValValAsnIleThrAlaArgArgLysLe 505

    591 C.................................................. 591
        :
    505 uThrValAsnSerSerIleSerIleGluArgGlySerHisLeuIleLeuH 522

    592 ........GGAGCAGGCAGACAATATTGGCGGTCTGATGAAGACGAACCC 633
                |||   |||   |||      :::   |||:::|||
    522 isSerGluGlyGlnGlyGlyGlnGlyValGlnIleAspLysAsp...... 536

    634 AATAACCGCGAAAGTTCATATCATATTGCAAGCGCATATTCTTGGCTCGT 683
                                 |||:::|||
    537 ........................IleThrSerGlu.............. 540

    684 CGGTGGCAATACCTTTGCACAAAATGGATCAGGTGGT............. 720
         |||||||||                   |||||||||
```

```
    541 .GlyGlyAsnLeuThrIleTyr......SerGlyGlyTrpValAspValH 555

    721 ..GGCACAGTCAACTTAGGTAGCGAAAAAATTAAACATAGCCCATATGGT 768
           :::::::::|||||||||                          |||
    555 isLysAsnIleThrLeuGlySer.......................Gly 563

    769 TTTTTACCAACAGGAGGCTCATTTGGCGAC..................AG 800
        ||||||             :::   ||||||                ||
    564 PheLeuAsnIleThrThrLysGluGlyAspIleAlaPheGluAspLysSe 580

    801 TGGCTCACCAATGTTTATCTATGATGCCCAAAAGCAAAAGTGGTTAATTA 850
        ||||         :::         ||||||
    580 rGlyArgAsnAsnLeuThrIleThrAlaGln................... 590

    851 ATGGGGTATTGCAAACAGGCAACCCCTATATAGGAAAAAGCAATGGCTTC 900
          |||:::::::   :::||||||                ||||||||||||
    591 ..GlyThrIleThrSerGlyAsn..............SerAsnGlyPhe 601

    901 CAGCTAGTTCGTAAAGATTGGTTCTATGATGAAATCTTTGCTGGAGATAC 950
        :::                        ::::::::::::    .  ::::::
    602 Arg.....................PheAsnAsnValSerLeuAsnSerLe 611

    951 CCATTCAGTATTCTACGAACCACATCAAAATGGGAAATACTTTTTTAACG 1000
                                        :::|||||||      |||:::|
    611 u..........................GlyGlyLysLeuSerPheThrA 619

   1001 ACAATAATAATGGCGCAGGAAAA.......................... 1023
        ||:::::::::      |||:::
    619 spSerArgGluAspArgGlyArgArgThrLysGlyAsnIleSerAsnLys 635

   1024 ....................................ATCGATGCCAAACA 1037
                                           |||:::   |||
    636 PheAspGlyThrLeuAsnIleSerGlyThrValAspIleSerMetLysAl 652

   1038 TAAACACTATTCTCTACCTTATAGATTAAAAACACGAACCGTT....... 1080
              |||      ||||||   |||   ||||||
    652 aProLysValSerTrpPheTyrArgAspLysGlyArgThrTyrTrpAsnV 669

   1081 ..........................CAATTGTTTAATGTTTCTTTATCC 1104
                                   :::   ||||||:::|||:::::
    669 alThrThrLeuAsnValThrSerGlySerLysPheAsnLeuSerIleAsp 685

   1105 GAGACAGCAAGAGAACCTGTTTATCATGCTGCAGGTGGGGTCAACAGTTA 1154
        :::|||                              |||   :::|||
    686 SerThr............................GlySerGlySerTh 692

   1155 TCGACCCAGACTGAATAATGGAGAAAATATTTCCTTTATTGACAAAGGAA 1204
           |||   ::::::|||:::|||   ::::::   |||      :::|
    692 rGlyProSerIleArgAsnAlaGlu...LeuAsnGlyIleThrPheAsnL 708

   1205 AAGGTGAATTGATACTTACCAGCAACATCAACCAAGGCGCGGGCGGTTTG 1254
        ||:::            |||   ||||||   ||||||:::
    708 ysAla............ThrPheAsnIleAlaGlnGlySer......... 717

   1255 TATTTTGAGGGTAATTTTACGGTCTCGCCT..............AAAAA 1289
                ::::||||||:::::::::                  |||::
    718 ......ThrAlaAsnPheSerIleLysAlaSerIleMetProPheLysSe 732

   1290 CAACGAAACGTGGCAA......GGCGCGGGCGTTCATATCAGTGATGGCA 1333
        :|||   ::::::        :::      :::   :::|||   |||:
    732 rAsnAlaAsnTyrAlaLeuPheAsnGluAspIleSerValSerGlyGlyG 749

   1334 GTACCGTTACTTGGAAAGTAAACGGC....................... 1359
        :::::|||::::::|||:::|||:::
    749 lySerValAsnPheLysLeuAsnAlaSerSerSerAsnIleGlnThrPro 765

   1360 .........GTGGCAAACGACCGCCTGTCCAAAATCGGCAAAGGCACGCT 1400
                 :::   ::::::::::::::      |||   :::|||||
    766 GlyValIleIleLysSerGlnAsnPheAsnValSerGlyGlySerThrLe 782

   1401 GCTGGTTCAAGCCAAAGGG.....................GAAAACC 1426
        |   ::::::|||:::|||                        ||||||:
    782 uAsnLeuLysAlaGluGlySerThrGluThrAlaPheSerIleGluAsnA 799
```

```
1427 AAGGCTCGGTCAGCGTGGGCGACGGTAAAGTCATCTTAGATCAG...CAG 1473
799 spLeuAsnLeuAsnAlaThrGlyGlyAsnIleThrIleArgGlnValGlu 815

1474 GCGGACGATCAA..................................... 1485
816 GlyThrAspSerArgValAsnLysGlyValAlaAlaLysLysAsnIleTh 832

1486 .....GGCAAAAAACAAGCCTTTAGTGAAA 1510
832 rPheLysGlyGlyAsnIleThrPheGlySerGlnLysAlaThrThrGluI 849

1511 TCGGCTTGGTCAGCGGCAGGGGGACGGTGCAACTGAATGCCGATAATCAG 1560
849 le.............LysGlyAsnValThrIleAsnLysAsnThrAsn 860

1561 TTCAACCCCGACAAACTCTATTTCGGCTTTCGCGGCGGACGTTTGGATTT 1610
861 AlaThr.....................LeuArgGlyAlaAsnPheAlaGl 870

1611 GAACGGGCATTCGCTTTCGTTCCACCGCATTCAAAATACCGATGAAGGGG 1660
870 uAsnLysSerProLeuAsnIle.....................AlaGlyA 880

1661 CGATGATTGTCAACCACAAT...CAAGACAAAGAATCCACCGTTACCATT 1707
880 snValIleAsnAsnGlyAsnLeuThrThrAlaGlySerIleIleAsnIle 896

1708 ACAGGCAATAAAGATATTACTACAACCGGCAAT................. 1740
897 AlaGlyAsnLeuThrValSerLysGlyAlaAsnLeuGlnAlaIleThrAs 913

1741 ..........AACAACAACTTGGATAGCAAAA 1762
913 nTyrThrPheAsnValAlaGlySerPheAspAsnAsnGlyAlaSerAsnI 930

1763 AAGAAATTGCCTACAACGGTTGGTTTGGCGAGAAAGATGCAACCAAAACG 1812
930 leSerIleAla...ArgGlyGlyAlaLysPheLysAspIleAsnAsnThr 945

1813 AACGGGCGGCTCAATCTGAATTACCAACCGGAAGAAGCGGATCGCACTTT 1862
946 SerSer...LeuAsnIleThrThrAsnSerAspThrThrTyrArgThrIl 961

1863 ACTGCTTTCCGGCGGAACAAATTTAAACGGCAATATCACGCAAACAAACG 1912
961 e.....................IleLysGlyAsnIleSerAsnLysSerG 971

1913 GCAAACTGTTTTTCAGCGGCAGACCGACACCGCACGCCTACAATCATTTA 1962
971 lyAspLeuAsnIleIleAspLysLysSer...AspAlaGluIleGlnIle 986

1963 GGAAGCGGGTGGTCAAAAATGGAAGGTATCCCACAAGGAGAAATCGTGTG 2012
987 GlyGlyAsnIleSerGlnLysGlu............GlyAsnLeuThrIl 999

2013 GGACAACGATTGGATCGAC............CGCACATTTAAAGCGGAAA 2050
999 eSerSerAspLysValAsnIleThrAsnGlnIleThrIleLysAla.... 1014

2051 ACTTCCATATTCAGGGCGGACAAGCG....................... 2076
1015 .....GlyValGluGlyGlyArgSerAspSerSerGluAlaGluAsnAla 1029

2077 ......GTGGTTTCCCGCAATGTTGCCAAAGTGGAAGGCGATTGGCATTT 2120
1030 AsnLeuThrIleGlnThrLysGluLeuLysLeuAlaGlyAspLeuAsnIl 1046

2121 AAGCAATCACGCCCAAGCAGTTTTCGGTGTCGCACCGCATCAAAGCCACA 2170
1046 eSerGlyPheAsnLysAlaGluIleThrAlaLysAsnGlySerAspLeuT 1063

2171 CAATCTGTACACGTTCGGACTGGACGGGTCTGACAAGTTGTACCGAAAAA 2220
1063 hrIleGlyAsnAlaSerGly............GlyAsnAlaAspAlaLys 1075

2221 ACCATTACCGACGATAAAGTGATTGCTTCATTGAGCAAGACCGACATCAG 2270
1076 LysValThrPheAspLysValLysAspSerLysIleSerThrAsp...Gl 1091

2271 AGGCAATGTCAGCCTTGCCGATCACGCTCATTTAAATCTCACAGGACTTG 2320
1091 yHisAsnValThrLeuAsnSerGlu.....................ValL 1101

2321 CCACACTCAACGGCAATCTTAGTGCAGGCGGAGAC......ACGCACTAT 2364
1101 ysThrSerAsnGlySerSerAsnAlaGlyAsnAspAsnSerThrGlyLeu 1117

2365 ACGGTTACG...CGCAACGCCACCCAAAACGGCAACCTCAGCCTCGTGGG 2411
1118 ThrIleSerAlaLysAspValThrValAsnAsnAsnValThr........ 1131

2412 CAATGCCCAAGCAACATTTAATCAAGCCACATTAAACGGCAACACATCGG 2461
1132 ....SerHisLysThrIleAsnIleSerAlaAlaAlaGlyAsnValThrT 1147

2462 CTTCGGACAATGCTTCATTTAATCTAAGCAACAACGCCGTA......... 2502
1147 hrLysGluGlyThrThrIleAsnAlaThrThrGlySerValGluValThr 1163

2503 ..................................CAAAACGGCAGTCT 2516
1164 AlaGlnAsnGlyThrIleLysGlyAsnIleThrSerGlnAsnValThrVa 1180

2517 GACGCTTTCCGACAAC......GCTAAGGCAAACGTAAGCCATTCCGCAC 2560
1180 lThrAlaThrGluAsnLeuValThrThrGluAsnAlaValIleAsnAlaT 1197

2561 TCAACGGCAATGTCTCCCTAGCCGATAAGGCAGTATTCCATTTTGAAAAC 2610
1197 hrSerGlyThrValAsnIleSerThrLys..................... 1206

2611 AGCCGCTTTACCGGAAAAATCAGCGGCGGCAAGGATACGGCA........ 2652
1207 .........ThrGlyAspIleLysGlyGlyIleGluSerThrSerGlyAs 1220

2653 .TTACACTTAAAAGACAGCGAATGGACGCTGCCGTCGGGCACGGAATTAG 2701
1220 nValAsnIleThrAlaSerGlyAsnThrLeu............LysValS 1233

2702 GCAATTTAAACCTTGACAACGCCACCATTACACTCAATTCCGCCTATCGA 2751
1233 erAsnIleThrGlyGlnAspValThrValThrAlaAsp............ 1245

2752 CACGATGCGGCAGGCGCGCAAACCGGCAGTGCGGCAGATGCGCCGCGCCG 2801
1246 .........AlaGlyAlaLeuThrThrThrAlaGly.............. 1254

2802 CCGTTCGCGCCGTTCCCTATTATCCGTTACGCCGCCAACTTCGGCAGAAT 2851
1255 .............SerThrIleSerAlaThrThrGlyAsnAla....... 1264

2852 CCCGTTTCAACACGCTGACGGTAAACGGCAAATTGAACGGTCAGGGAACA 2901
1265 ........AsnIleThrThrLysThrGlyAspIleAsnGlyLys...... 1276

2902 TTCCGCTTTATGTCGGAACTCTTCGGCTACCGCAGCGGCAAATTGAAGCT 2951

1276 ..................................................... 1276

2952 GGCGGAAAGTTCCGAAGGCACTTAC.....................ACCT 2980
1277 .ValGluSerSerSerGlySerValThrLeuValAlaThrGlyAlaThrL 1293

2981 TGGCTGTCAACAATACCGGCAACGAACCCGTAAGTCTCGAGCAATTGACG 3030
```

```
     |||||||:::|||   :::::::::   |||:::         :::|||
1293 euAlaValGlyAsnIleSerGlyAsnThrValThr.........IleThr 1306

3031 GTAGTGGAAGGAAAAGACAACACACCGCTGTCCGAAAATCTTAATTTCAC 3080
     :::   :::|||||||   ::::::   ::::::::::::::|||   ||
1307 AlaAspSerGlyLysLeuThrSerThrValGlySerThrIleAsnGlyTh 1323

3081 CCTGCAAAACGAACACGTCGATGCCGGCGCATGGCGTTATCAGCTTATCC 3130
     |
1323 r.................................................. 1323

3131 GCAAAGACGGCGAGTTCCGCCTGCATAATCCGGTCAAAGAACAAGAGCTT 3180
                        ::::::   :::   :::::::::
1324 .......................AsnSerValThrThrSerSerGlnSer 1332

3181 TCCGACAAACTCGGCAAGGCGGGAGAAACAGAGGCCGCCTTGACGGCAAA 3230
     :::|||      |||      :::   :::   :::   :::|||||||::
1333 GlyAspIleGluGlyThrIleSerGlyAsnThrValAsnValThrAlaSe 1349

3231 ACAGGCACAACTT............GCCGCCAAACAACAGGCGGAAAAAG 3268
     :   ::::::|||            :::||||||   :::|||   |||:
1349 rThrGlyAspLeuThrIleGlyAsnSerAlaLysValGluAla...LysA 1365

3269 ACAACGCGCAAAGCCTTGACGCGCTGATTGCGGCCGGGCGCAATGCCACC 3318
     :::::|||   :::|||   |||          :::|||:::   :::|||
1365 snGlyAlaAlaThrLeuThrAla......GluSerGlyLysLeuThrThr 1379

3319 GAAAAGGCAGAAAGTGTT............................GC 3338
     :::   ::::::|||:::                             ||
1380 GlnThrGlySerSerIleThrSerSerAsnGlyGlnThrThrLeuThrAl 1396

3339 CGAACCGGCCCGGCAGGCAGGCGGGGAAAATGCC.............. 3372
     |:::   :::      ||||||:::   ||||||
1396 aLysAspSerSerIleAlaGlyAsnIleAsnAlaAlaAsnValThrLeuA 1413

3373 ........GGCATTATGCAGGCGGAGGAAGAGAAAAAACGGGTGCAGGCG 3414
             |||   :::   :::      ::::::|||   :::::::|||
1413 snThrThrGlyThrLeuThrThrThrGlyAspSerLys...IleAsnAla 1428

3415 GATAAAGACACCGCCTTGGCGAAACAGCGCGAAGCGGAAACCCGGCCGGC 3464
        :::   |||        :::   ::::::|||:::        ||
1429 ThrSerGlyThrLeuThrIleAsnAlaLysAspAlaLysLeuAspGlyAl 1445

3465 TACCACCGCCTTCCCCCGCGCCCGCCGCGCCCGCCGGGATTTGCCGCAAC 3514
     |:::::::::
1445 aAlaSerGly..................................... 1448

3515 CGCAGCCCCAACCGCAACCCCAACCGCAGCGCGACCTGATCAGCCGTTAT 3564
                              :::|||   :::::::::
1449 .........................AspArgThrValValAsnAlaThr 1456

3565 GCCAATAGCGGTTTGAGTGAATTTTCCGCCACGCTCAACAGCGTTTTCGC 3614
           ||||||   ::::::   :::|||      :::|||
1457 AsnAlaSerGlySerGlyAsnValThrAlaLysThrSerSerSerValAs 1473

3615 CGTACAGGACGAATTGGAC..............CGCGTGTTTGCCGAAG 3649
       :::      :::|||:::               :::::::::::|||:
1473 nIleThrGlyAspLeuAsnThrIleAsnGlyLeuAsnIleIleSerGluA 1490

3650 ACCGCCGCAACGCCGTTTGGACAAGCGGCATCCGG...GACACCAAACAC 3696
     ::   ||||||:::|||         |||   :::   |||:::|||:::
1490 snGlyArgAsnThrValArgLeuArgGlyLysGluIleAspValLysTyr 1506

3697 TACCGTTCGCAAGATTTCCGCGCCTACCGCCAACAAACCGACCTGCGCCA 3746
             |||    ::::::   ::::::::::::        ||||
1507 Ile......GlnPro.GlyValAlaSerValGluGluValIleGluAlaL 1521

3747 AATCGGTATGCAGAAAAACC.........................TCG 3769
     ||   |||   :::|||:::                          :::
1521 ysArgValLeuGluLysValLysAspLeuSerAspGluGluArgGluThr 1537

3770 GCAGCGGGCGCGTCGGCATCCTGTTTTCGCACAACCGGACCGGAAACACC 3819
        |||      ::::::|||:::   |||   :::   |||::::::
```

```
     1538 LeuAlaLysLeuGlyValSerAlaValArgPheValGluProAsnAsnAl 1554

     3820 TTCGACGACGGCATCGGCAACTCGGCACGGCTTGCCCACGGTGCCGTTTT 3869
               |||:::         |||::::::   :::   |||   |||   |
     1554 aIleThrVal......AsnThrGlnAsnGluPheThrThrLysProSerS 1569

     3870 CGGGCAATACGGCATCGGCAGGTTCGACA......TCGGCATCAGCGCGG 3913
          ||       |||   |||   |||::::::      |||::::::
     1569 erGlnValThrIleSerGluGlyLysAlaCysPheSerSerGlyAsnGly 1585

     3914 GCGCGGGTTTTAGTAGCGGCAGCC 3937
          |||||||||   :::   :::|||
     1586 AlaArgValCysThrAsnValAla 1593

seq_name: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:US-08-617-697-10

seq_documentation_block:
; Sequence 10, Application US/08617697
; Patent No. 5977336
;  GENERAL INFORMATION:
;    APPLICANT:  Barenkamp, Stephen J
;    TITLE OF INVENTION:  High Molecular Weight Surface Proteins
;    TITLE OF INVENTION:  of No. 5977336-Typeable Haemophilus
;    NUMBER OF SEQUENCES:  11
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Shoemaker and Mattare, Ltd.
;      STREET:  2001 Jefferson Davis Hwy., 1203 Crystal Plaza
;      STREET:  Bldg. 1
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22202-0286
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/617,697
;      FILING DATE:  01-APR-1996
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/302,832
;      FILING DATE:  05-OCT-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US PCT/US93/02166
;      FILING DATE:  16-MAR-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Berkstresser, Jerry W
;      REGISTRATION NUMBER:  22,651
;      REFERENCE/DOCKET NUMBER:  1038-557
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 415-0810
;      TELEFAX:  (703) 415-0813
;  INFORMATION FOR SEQ ID NO:  10:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1600 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-617-697-10


alignment_scores:
            Quality:  301.50              Length:   1508
              Ratio:  0.433                 Gaps:     71
 Percent Similarity:  46.220   Percent Identity: 19.761

alignment_block:
 US-09-303-518D-653 x US-08-617-697-10   ..

 Align seg 1/1  to: US-08-617-697-10  from: 1  to: 1600
```

```
178 AATAAAGGCAAGTTTGCAGTCGGGGCGAAAGATATTGAGGTTTACAACAA 227
199 AsnHisGlyLeuIleThrValGly...LysAspGlySerVal........ 211

228 AAAAGGGGAGTTGGTCGGCAAATCGATGACGAAAGCCCCGATGATTGATT 277
212 .......AsnLeuIleGlyGlyLysValLysAsnGlu.............. 221

278 TTTCTGTGGTATCGCGTAACGGC......GTGGCGGCATTGGCGGGCGAT 321
222 ..GlyValIleSerValAsnGlyGlySerIleSerLeuLeuAlaGlyGln 237

322 CAA.....................................TATATTGTGAG 335
238 LysIleThrIleSerAspIleIleAsnProThrIleThrTyrSerIleAl 254

336 CGTGGCACATAACGGCGGCTATAACAATGTTGATTTTGGTGCGGAGGGAA 385
254 aAlaProGluAsnGluAlaIleAsnLeuGlyAspIlePheAlaLysGlyG 271

386 GCAATCCCGATCAGCACCGCTTTTCTTACCAAATTGTGAAAAGAAATAAT 435
271 lyAsn.......................................... 272

436 TATAAAGCAGGGACTAACGGCCATCCTTATGGCGGCGATTATCATATGCC 485

272 ..............................................  272

486 GCGTTTGCACAAATTTGTCACAGATGCAGAACCTGTTGAGATGACCAGTT 535

272 ..............................................  272

536 ATATGGATGGGTGGAAATACGCTGATTTAAATAAATACCCTGATCGTGTT 585
273 .........................IleAsn.............Val 275

586 CGAATCGGAGCAGGCAGACAATATTGGCGGTCTGATGAAGACGAACCCAA 635
276 ArgAlaAlaThrIleArgAsnLysGlyLysLeuSerAlaAspSerValSe 292

636 TAACCGCGAAAGTTCATATCATATTGCAAGCGCATATTCTTGG......C 679
292 rLysAspLysSerGlyAsnIleValLeuSerAlaLysGluGlyGluAlaG 309

680 TCGTCGGTGGCAATACCTTTGCACAAAATGGATCAGGT...GGTGGCACA 726
309 luIleGlyGlyValIleSerAlaGlnAsnGlnGlnAlaLysGlyGlyLys 325

727 GTCAACTTAGGTAGCGAAAAAATTAAACATAGCCCATATGGTTTTTTACC 776
326 LeuMetIleThrGlyAspLysValThr.................LeuLy 336

777 AACAGGAGGCTCATTTGGCGACAGTGGCTCACCAATGTTTATCTATGATG 826
336 sThrGlyAlaValIleAspLeuSerGlyLys................... 346

827 CCCAAAAGCAAAAGTGGTTAATTAATGGGGTATTGCAAACAGGCAACCCC 876
347 .................................GluGlyGlyGluThr 351

877 TATATAGGA.................AAAAGCAATGGCTTCCAGCTAGT 908
352 TyrLeuGlyGlyAspGluArgGlyGluGlyLysAsnGlyIleGlnLeuAl 368

909 TCGTAAAGATTGGTTCTATGATGAAATCTTTGCTGGAGATACCCATTCAG 958
368 aLysLys............ThrThrLeuGluLysGlySerThrIleAsnV 381

959 TATTCTACGAACCACATCAAAATGGGAAA..................... 987
381 al.................SerGlyLysGluLysGlyGlyArgAlaIle 391

988 ..............TACTTTTTTAACGACAATAATAATGGCGCAGGAAA 1022
```

```
392 ValTrpGlyAspIleAlaLeuIleAspGlyAsnIleAsnAlaGlnGlySe 408

1023 AATCGATGCCAAACATAAACACTATTCTCTACCTTATAGATTAAAAACAC 1072
408 rAspIleAlaLysThrGlyGlyPhe..............ValGluThrS 420

1073 GAACCGTTCAATTG..............TTTAATGTTTCTTTATCCGAG 1107
420 erGlyHisAspLeuSerIleGlyAspAspValIleValAspAlaLysGlu 436

1108 ACAGCAAGAGAACCT............GTTTATCATGCTGCAGGTGGGGT 1145
437 TrpLeuLeuAspProAspAspValSerIleGluThrLeuThrSerGlyAr 453

1146 CAACAGTTATCGACCCAGACTGAATAATGGAGAAAATATTTCCTTTATT. 1194
453 gAsn..................AsnThrGlyGluAsnGlnGlyTyrThrT 464

1195 .................GACAAAGGAAAAGGTGAATTGATACTTACCAGC 1227
464 hrGlyAspGlyThrLysGluSerProLysGlyAsnSerIleSerLysPro 480

1228 AACATC...AACCAAGGCGCGGGCGGTTTGTATTTTGAGGGTAATTTTAC 1274
481 ThrLeuThrAsnSerThrLeuGluGlnIleLeuArgArgGlySerTyrVa 497

1275 GGTCTCGCCTAAAAACAACGAAACGTGGCAAGGCGCGGGCGTTCATATCA 1324
497 lAsnIleThrAlaAsnAsnArgIleTyrValAsnSerSerIleAsnLeuS 514

1325 GTGATGGCAGT...ACCGTTACTTGGAAAGTAAACGGCGTG...GCAAAC 1368
514 erAsnGlySerLeuThrLeuHisThrLysArgAspGlyValLysIleAsn 530

1369 GACCGCCTGTCCAAAATCGGCAAAGGCACGCTGCTGGTTCAAGCCAAAGG 1418
531 GlyAspIleThrSerAsnGluAsnGlyAsnLeuThrIleLysAlaGlySe 547

1419 G......GAAAACCAAGGCTCGGTCAGCGTGGGCGACGGTAAAGTCATCT 1462
547 rTrpValAspValHisLysAsnIleThrLeuGlyThrGly......PheL 562

1463 TAGATCAGCAGGCGGACGAT..................CAAGGCAAAAAA 1494
562 euAsnIleValAlaGlyAspSerValAlaPheGluArgGluGlyAspLys 578

1495 CAAGCCTTTAGTGAAATCGGCTTGGTCAGCGGCAGGGGGACGGTGCAACT 1544
579 AlaArgAsnAlaThrAspAlaGlnIleThrAlaGlnGlyThrIleThrVa 595

1545 GAATGCCGATAATCAGTTCAACCCCGACAAACTCTATTTCGGCTTTCGCG 1594
595 lAsnLysAspAspLys..................GlnPheArgPheAsnA 606

1595 GCGGACGTTTGGATTTGAACGGGCATTCGCTTTCGTTCCACCGCATTCAA 1644
606 snValSerLeuAsnGlyThrGlyLysGlyLeuLysPheIleAlaAsnGln 622

1645 AATACCGATGAAGGGGCGATGATTGTCAACCACAATCAAGACAAAGAATC 1694
623 AsnAsn..................PheThrHisLysPheAspGlyGlu.. 632

1695 CACCGTTACCATTACAGGCAATAAAGATATTACTACAACCGGCAATAACA 1744
633 ....IleAsnIleSerGlyIleValThrIleAsnGlnThr.......... 644

1745 ACAACTTGGATAGCAAAAAAGAAATTGCCTACAACGGTTGGTTTGGCGAG 1794
645 ...........ThrLysLysAspValLysTyr......TrpAsnAlaSer 655

1795 AAAGATGCAACCAAAACGAACGGGCGGCTCAATCTGAATTACCAACCGGA 1844
```

```
 3397 .....AAAAAACGGGTGCAGGCGGATAAAG 3421
                                 |||   ::::::   :::
 1125 lThrValAsnAsnAsnValThrSerHisLysThrIleAsnIleSerAlaA 1142

 3422 ACACCGCCTTGGCGAAACAGCGCGAAGCGGAAACCCGGCCGGCTACCACC 3471
        ::::::   :::      :::|||:::   |||      |||||||||
 1142 laAlaGlyAsnValThrThrLysGluGlyThrThrIleAsnAlaThrThr 1158

 3472 GCCTTCCCCCGCGCCCGCCGCGCCCGCCGGGATTTGCCGCAACCGCAGCC 3521
      ::: |||         :::|||      |||      :::::::::
 1159 Gly.SerValGluValThrAlaGlnAsnGlyThrIleLysGlyAsnIleT 1175

 3522 CCAACCGCAACCCCAACCGCAGCGCGACCTGATCAGCCGTTATGCCAATA 3571
        :::::::|||   :::   :::|||||||   :::   |||
 1175 hrSerGlnAsnValThrValThrAlaThrGluAsnLeuValThrThrGlu 1191

 3572 GCGGTTTGAGTGAATTTTCCGCCACGCTCAACAGCGTTTTCGCCGTACAG 3621
        :::   :::|||            ::::::   :::|||
 1192 AsnAlaValIleAsnAlaThrSerGlyThrValAsnIleSer..... 1205

 3622 GACGAATTGGACCGCGTGTTTGCCGAAGACCGCCGCAACGCCGTTTGGA. 3670
                               ||||||:::            |||
 1206 .....ThrLysThrGlyAspIleLysGlyGlyI 1215

 3671 ..CAAGCGGCATCCGGGACACCAAACACTACCGTTCGC..... 3706
        :::::::::||||||:::   |||   |||:::
 1215 leGluSerThrSerGlyAsnValAsnIleThrAlaSerGlyAsnThrLeu 1231

 3707 AAGATTTCCGCGCCTACCGCCAACAAACCGACCTGCGCCAAATCGGTATG 3756
      |||:::|||      |||::::::      |||   :::   ::::::
 1232 LysValSerAsnIleThrGlyGlnAspValThrValThrAlaAspAlaGl 1248

 3757 CAGAAAAACCTCGGCAGCGGGCGCGTCGGCATCCTGTTTTCGCACAACCG 3806
             |||:::::|||||:::::   |||:::      |||||||
 1248 yAlaLeuThrThrThrAlaGlySerThrIleSerAla......ThrThrG 1263

 3807 GA..... 3808
      ||
 1263 lyAsnAlaAsnIleThrThrLysThrGlyAspIleAsnGlyLysValGlu 1279

 3809 CCGGAAACACCTTCGACGACGGCATCGGCAACTCGGCACGGCTTGCCCAC 3858
        ::::::   |||:::|||      ||||||         |||   ::
 1280 SerSerSerGlySerValThrLeuValAlaThrGlyAlaThrLeuAlaVa 1296

 3859 GGTGCCGTTTTCGGGCAAT.....ACGGCATCGGCAGGTTCGA 3896
      :      :::|||||||||          ||||||::::::|||:::
 1296 lGlyAsnIleSerGlyAsnThrValThrIleThrAlaAspSerGlyLysL 1313

 3897 CATCGGCATCAGCGCGGGCGCGGGTTTTAGTAGCGGCAGCCTTTCAGACG 3946
        ::::::::::::   :::   :::      :::   :::      |||
 1313 euThrSerThrValGlySerThrIleAsnGlyThrAsnSerValThrThr 1329

 3947 GCATCAGAGGCAAAA.....TCCGCCGCCGCGTGCTG 3978
      :::|||::::::                |||:::   :::
 1330 SerSerGlnSerGlyAspIleGluGlyThrIleSerGlyAsnThrValAs 1346

 3979 CATTACGGCATTCAGGCAAGAT.....ACCGCGCAGGTTT 4013
       :::|||||||         |||              :::|||:::|||:
 1346 nValThrAlaSerThrGlyAspLeuThrIleGlyAsnSerAlaLysValG 1363

 4014 CGGCGGATTCGGCATCGAACCGCACATCGGCGCAACGCGCTATTTCGTCC 4063
      ::|||   :::::::::         :::|||:::             |||
 1363 luAlaLysAsnGlyAlaAlaThrLeuThrAlaGlu.....Ser 1375

 4064 AAAAAGCGGATTACCGATACGAAAACGTCAATATCGCCACCCCGGGCCTT 4113
        |||      |||:::|||   :::||||||:::      ::::::
 1376 GlyLysLeuThrThrGlnThrGlySerSerIleThrSerSerAsnGly.. 1391

 4114 GCATTCAACCGCTACCGCGCGG 4135
      ::::::|||   ||||||:::
 1392 .GlnThrThrLeuThrAlaLys 1398

seq_name: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:US-08-728-470-10
```

```
seq_documentation_block:
; Sequence 10, Application US/08728470
; Patent No. 5928651
;  GENERAL INFORMATION:
;    APPLICANT:  Barenkamp, Stephen J
;    TITLE OF INVENTION:  High Molecular Weight Surface Proteins
;    TITLE OF INVENTION:  of No. 5928651-Typeable Haemophilus
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Shoemaker and Mattare, Ltd.
;      STREET:  2001 Jefferson Davis Hwy., 1203 Crystal Plaza
;      STREET:  Bldg. 1
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22202-0286
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/728,470
;      FILING DATE:
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/302,832
;      FILING DATE:  16-MAR-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US PCT/US93/02166
;      FILING DATE:  16-MAR-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9205704.1
;      FILING DATE:  16-MAR-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Berkstresser, Jerry W
;      REGISTRATION NUMBER:  22,651
;      REFERENCE/DOCKET NUMBER:  1038-633
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 415-0810
;      TELEFAX:  (703) 415-0813
;  INFORMATION FOR SEQ ID NO:  10:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1529 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-728-470-10


alignment_scores:
            Quality:  292.50              Length:   1494
              Ratio:  0.415                 Gaps:     77
 Percent Similarity:  47.189   Percent Identity: 20.415

alignment_block:
 US-09-303-518D-653 x US-08-728-470-10   ..

 Align seg 1/1  to: US-08-728-470-10  from: 1  to: 1529

     178 AATAAAGGCAAGTTTGCAGTCGGGGCGAAAGATATTGAGGTTTACAACAA 227
         |||   |||   ::::::||||||   ||||||   :::|||
     127 AsnHisGlyLeuIleThrValGly...LysAspGlySerVal..... 139

     228 AAAAGGGGAGTTGGTCGGCAAATCGATGACGAAAGCCCCGATGATTGATT 277
               :::|||:::|||   ::::::   :::
     140 .......AsnLeuIleGlyGlyLysValLysAsnGlu..... 149

     278 TTTCTGTGGTATCGCGTAACGGC......GTGGCGGCATTGGCGGGCGAT 321
           :::|||:::|||   ||||||      ::::::   |||||||||:::
     150 ..GlyValIleSerValAsnGlyGlySerIleSerLeuLeuAlaGlyGln 165
```

```
 322 CAA.................................TATATTGTGAG 335
     :::                                |||   :::::
 166 LysIleThrIleSerAspIleIleAsnProThrIleThrTyrSerIleAl 182

 336 CGTGGCACATAACGGCGGCTATAACAATGTTGATTTTGGTGCGGAGGGAA 385
     ::::   :::|||   :::   |||     |||:::   |||::::|||:
 182 aAlaProGluAsnGluAlaIleAsnLeuGlyAspIlePheAlaLysGlyG 199

 386 GCAATCCCGATCAGCACCGCTTTTCTTACCAAATTGTGAAAAGAAATAAT 435
     ::|||
 199 lyAsn..................................................... 200

 436 TATAAAGCAGGGACTAACGGCCATCCTTATGGCGGCGATTATCATATGCC 485

 200 ..................................................... 200

 486 GCGTTTGCACAAATTTGTCACAGATGCAGAACCTGTTGAGATGACCAGTT 535

 200 ..................................................... 200

 536 ATATGGATGGGTGGAAATACGCTGATTTAAATAAATACCCTGATCGTGTT 585
                              :::|||               |||
 201 ........................IleAsn.............Val 203

 586 CGAATCGGAGCAGGCAGACAATATTGGCGGTCTGATGAAGACGAACCCAA 635
     |||   ::::::   |||:::     :::   :::   |||:::   ::
 204 ArgAlaAlaThrIleArgAsnLysGlyLysLeuSerAlaAspSerValSe 220

 636 TAACCGCGAAAGTTCATATCATATTGCAAGCGCATATTCTTGG......C 679
     ::::   :::|||:::      :::   ||||||   :::
 220 rLysAspLysSerGlyAsnIleValLeuSerAlaLysGluGlyGluAlaG 237

 680 TCGTCGGTGGCAATACCTTTGCACAAAATGGATCAGGT...GGTGGCACA 726
       :::||||||         |||||||||   ::::::   ||||||
 237 luIleGlyGlyValIleSerAlaGlnAsnGlnGlnAlaLysGlyGlyLys 253

 727 GTCAACTTAGGTAGCGAAAAAATTAAACATAGCCCATATGGTTTTTTACC 776
     :::   :::   ::::::|||:::                      |||
 254 LeuMetIleThrGlyAspLysValThr.................LeuLy 264

 777 AACAGGAGGCTCATTTGGCGACAGTGGCTCACCAATGTTTATCTATGATG 826
      ||||||:::   :::      ||||||:::
 264 sThrGlyAlaValIleAspLeuSerGlyLys..................... 274

 827 CCCAAAAGCAAAAGTGGTTAATTAATGGGGTATTGCAAACAGGCAACCCC 876
                                  :::   |||:::
 275 ................................GluGlyGlyGluThr 279

 877 TATATAGGA.................AAAAGCAATGGCTTCCAGCTAGT 908
     |||:::|||                :::|||||||:::|||||||::
 280 TyrLeuGlyGlyAspGluArgGlyGluGlyLysAsnGlyIleGlnLeuAl 296

 909 TCGTAAAGATTGGTTCTATGATGAAATCTTTGCTGGAGATACCCATTCAG 958
     ::::|||                  :::      |||:::|||   :::|
 296 aLysLys............ThrThrLeuGluLysGlySerThrIleAsnV 309

 959 TATTCTACGAACCACATCAAAATGGGAAA..................... 987
     ||                 :::||||||
 309 al.................SerGlyLysGluLysGlyGlyArgAlaIle 319

 988 ..............TACTTTTTTAACGACAATAATAATGGCGCAGGAAA 1022
                     :::::::::   |||   |||:::   |||::
 320 ValTrpGlyAspIleAlaLeuIleAspGlyAsnIleAsnAlaGlnGlySe 336

1023 AATCGATGCCAAACATAAACACTATTCTCTACCTTATAGATTAAAAACAC 1072
     :      ||||||         :::               ::::::|||
 336 rAspIleAlaLysThrGlyGlyPhe..............ValGluThrS 348

1073 GAACCGTTCAATTG..............TTTAATGTTTCTTTATCCGAG 1107
            :::|||                  |||:::   :::|||
 348 erGlyHisAspLeuSerIleGlyAspAspValIleValAspAlaLysGlu 364

1108 ACAGCAAGAGAACCT............GTTTATCATGCTGCAGGTGGGGT 1145
```

```
            :::|||          :::         ::::::|||
 365 TrpLeuLeuAspProAspAspValSerIleGluThrLeuThrSerGlyAr 381

1146 CAACAGTTATCGACCCAGACTGAATAATGGAGAAAATATTTCCTTTATT. 1194
      |||                 |||:::|||||||||   ::::::
 381 gAsn.................AsnThrGlyGluAsnGlnGlyTyrThrT 392

1195 .................GACAAAGGAAAAGGTGAATTGATACTTACCAGC 1227
                       ::::::   ||||||:::   |||
 392 hrGlyAspGlyThrLysGluSerProLysGlyAsnSerIleSerLysPro 408

1228 AACATC...AACCAAGGCGCGGGCGGTTTGTATTTTGAGGGTAATTTTAC 1274
     ::::::   |||:::          :::      :::|||::::::::
 409 ThrLeuThrAsnSerThrLeuGluGlnIleLeuArgArgGlySerTyrVa 425

1275 GGTCTCGCCTAAAAACAACGAAACGTGGCAAGGCGCGGGCGTTCATATCA 1324
     :           ||||||:::   :::   ::::::::::::::::::|
 425 lAsnIleThrAlaAsnAsnArgIleTyrValAsnSerSerIleAsnLeuS 442

1325 GTGATGGCAGT...ACCGTTACTTGGAAAGTAAACGGCGTG...GCAAAC 1368
     ||:::||||||   |||:::      |||   :::||||||      |||
 442 erAsnGlySerLeuThrLeuHisThrLysArgAspGlyValLysIleAsn 458

1369 GACCGCCTGTCCAAAATCGGCAAAGGCACGCTGCTGGTTCAAGCCAAAGG 1418
          :::::::::      :::|||:::|||   ::::::|||   ::
 459 GlyAspIleThrSerAsnGluAsnGlyAsnLeuThrIleLysAlaGlySe 475

1419 G......GAAAACCAAGGCTCGGTCAGCGTGGGCGACGGTAAAGTCATCT 1462
     :      :::   :::   :::::::::::::|||   |||      :::|
 475 rTrpValAspValHisLysAsnIleThrLeuGlyThrGly......PheL 490

1463 TAGATCAGCAGGCGGACGAT.................CAAGGCAAAAAA 1494
     ||:::      |||   |||                  :::|||   |||
 490 euAsnIleValAlaGlyAspSerValAlaPheGluArgGluGlyAspLys 506

1495 CAAGCCTTTAGTGAAATCGGCTTGGTCAGCGGCAGGGGGACGGTGCAACT 1544
               :::      :::   :::::::::::|||||:::   ::
 507 AlaArgAsnAlaThrAspAlaGlnIleThrAlaGlnGlyThrIleThrVa 523

1545 GAATGCCGATAATCAGTTCAACCCCGACAAACTCTATTTCGGCTTTCGCG 1594
     :|||   |||::::::                      |||   |||::::
 523 lAsnLysAspAspLys.................GlnPheArgPheAsnA 534

1595 GCGGACGTTTGGATTTGAACGGGCATTCGCTTTCGTTCCACCGCATTCAA 1644
     ::      ::::::   :::|||   :::|||:::|||         |||
 534 snValSerIleAsnGlyThrGlyLysGlyLeuLysPheIleAlaAsnGln 550

1645 AATACCGATGAAGGGGCGATGATTGTCAACCACAATCAAGACAAAGAATC 1694
     |||:::                  :::|||:::   |||   |||
 551 AsnAsn.................PheThrHisLysPheAspGlyGlu.. 560

1695 CACCGTTACCATTACAGGCAATAAAGATATTACTACAACCGGCAATAACA 1744
         ::::::|||:::|||         |||:::   |||
 561 ....LeuAsnIleSerGlyIleValThrIleAsnGlnThr.......... 572

1745 ACAACTTGGATAGCAAAAAAGAAATTGCCTACAACGGTTGGTTTGGCGAG 1794
               :::||||||::::::   |||      |||   ::::::
 573 ...........ThrLysLysAspValLysTyr......TrpAsnAlaSer 583

1795 AAAGATGCAACCAAAACGAACGGGCGGCTCAATCTG.............. 1830
     ||||||:::      :::   :::   |||:::|||
 584 LysAspSerTyrTrpAsnValSerSerLeuThrLeuAsnThrValGlnLy 600

1831 ................................AATTACCAACCGGAAG 1846
                                     |||   |||      :
 600 sPheThrPheIleLysPheValAspSerGlySerAsnSerGlnAspLeuA 617

1847 AAGCGGATCGCACTTTACTG.....................CTTTCCGGC 1875
     :::::::::|||      :::                  ::::::|||
 617 rgSerSerArgArgSerPheAlaGlyValHisPheAsnGlyIleGlyGly 633

1876 GGAACAAATTTAAACGGCAATATCACGCAAACAAACGGCAAACTGTTTTT 1925
       ||||||:::|||                ::::|||:::|||   :::||
```

```
   3564 TGCCAATAGCGGTTTGAGTGAATTTTCCGCCACGCTCAACAGCGTTTTCG 3613
                  :::   :::|||           ::::::   :::|||
   1119 hrThrGluAsnAlaValIleAsnAlaThrSerGlyThrValAsnIleSer 1135

   3614 CCGTACAGGACGAATTGGACCGCGTGTTTGCCGAAGACCGCCGCAACGCC 3663
                                      ||||||:::
   1136 ............................ThrLysThrGlyAspIleLy 1142

   3664 GTTTGGA...CAAGCGGCATCCGGGACACCAAACACTACCGTTCGC.... 3706
           |||   :::::::::||||||:::   |||   |||:::
   1142 sGlyGlyIleGluSerThrSerGlyAsnValAsnIleThrAlaSerGlyA 1159

   3707 ........AAGATTTCCGCGCCTACCGCCAACAAACCGACCTGCGCCAAA 3748
                ||||:::|||      |||::::::      |||   :::
   1159 snThrLeuLysValSerAsnIleThrGlyGlnAspValThrValThrAla 1175

   3749 TCGGTATGCAGAAAAACCTCGGCAGCGGGCGCGTCGGCATCCTGTTTTCG 3798
        ::::::         |||:::::::||||||:::::   |||:::
   1176 AspAlaGlyAlaLeuThrThrThrAlaGlySerThrIleSerAla..... 1190

   3799 CACAACCGGA..................................... 3808
         |||||||||
   1191 .ThrThrGlyAsnAlaAsnIleThrThrLysThrGlyAspIleAsnGlyL 1207

   3809 ........CCGGAAACACCTTCGACGACGGCATCGGCAACTCGGCACGGC 3850
                  ::::::   |||:::|||      ||||||
   1207 ysValGluSerSerSerGlySerValThrLeuValAlaThrGlyAlaThr 1223

   3851 TTGCCCACGGTGCCGTTTTCGGGCAAT............ACGGCATCGGC 3888
        |||   :::      :::|||||||||            ||||||:::::
   1224 LeuAlaValGlyAsnIleSerGlyAsnThrValThrIleThrAlaAspSe 1240

   3889 AGGTTCGACATCGGCATCAGCGCGGGCGCGGGTTTTAGTAGCGGCAGCCT 3938
        :|||:::   :::::::::::   :::   :::      :::   :::
   1240 rGlyLysLeuThrSerThrValGlySerThrIleAsnGlyThrAsnSerV 1257

   3939 TTCAGACGGCATCAGAGGCAAAA.................TCCGCCGCC 3970
              |||:::|||::::::                  |||:::
   1257 alThrThrSerSerGlnSerGlyAspIleGluGlyThrIleSerGlyAsn 1273

   3971 GCGTGCTGCATTACGGCATTCAGGCAAGAT..............ACCGC 4005
        :::      :::||||||         |||               :::||
   1274 ThrValAsnValThrAlaSerThrGlyAspLeuThrIleGlyAsnSerAl 1290

   4006 GCAGGTTTCGGCGGATTCGGCATCGAACCGCACATCGGCGCAACGCGCTA 4055
        |:::|||:::|||   :::::::::         :::|||:::
   1290 aLysValGluAlaLysAsnGlyAlaAlaThrLeuThrAlaGlu....... 1304

   4056 TTTCGTCCAAAAAGCGGATTACCGATACGAAAACGTCAATATCGCCACCC 4105
              |||   |||      |||:::|||   :::||||||:::
   1305 .....SerGlyLysLeuThrThrGlnThrGlySerSerIleThrSerSer 1319

   4106 CGGGCCTTGCATTCAACCGCTACCGCGCGG 4135
        ::::::   ::::::|||   |||||||:::
   1320 AsnGly...GlnThrThrLeuThrAlaLys 1328

seq_name: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-08-719-641-10

seq_documentation_block:
; Sequence 10, Application US/08719641
; Patent No. 6218141
;  GENERAL INFORMATION:
;    APPLICANT:  Barenkamp, Stephen J
;    TITLE OF INVENTION:  High Molecular Weight Surface Proteins
;    TITLE OF INVENTION:  of No. 6218141-Typeable Haemophilus
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Shoemaker and Mattare, Ltd.
;      STREET:  2001 Jefferson Davis Hwy., 1203 Crystal Plaza
;      STREET:  Bldg. 1
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22202-0286
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/719,641
;      FILING DATE:
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/302,832
;      FILING DATE:  16-SEP-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US PCT/US93/02166
;      FILING DATE:  16-MAR-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9205704.1
;      FILING DATE:  16-MAR-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Berkstresser, Jerry W
;      REGISTRATION NUMBER:  22,651
;      REFERENCE/DOCKET NUMBER:  1038-625
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 415-0810
;      TELEFAX:  (703) 415-0813
;  INFORMATION FOR SEQ ID NO:  10:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1529 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-719-641-10


alignment_scores:
            Quality:  292.50              Length:   1494
              Ratio:  0.415                 Gaps:     77
 Percent Similarity:  47.189   Percent Identity: 20.415

alignment_block:
 US-09-303-518D-653 x US-08-719-641-10   ..

 Align seg 1/1  to: US-08-719-641-10  from: 1  to: 1529

    178 AATAAAGGCAAGTTTGCAGTCGGGGCGAAAGATATTGAGGTTTACAACAA 227
        |||   |||   ::::::||||||   ||||||   ::|||
    127 AsnHisGlyLeuIleThrValGly...LysAspGlySerVal........ 139

    228 AAAAGGGGAGTTGGTCGGCAAATCGATGACGAAAGCCCCGATGATTGATT 277
              :::|||:::|||   ::::::   :::
    140 .......AsnLeuIleGlyGlyLysValLysAsnGlu............. 149

    278 TTTCTGTGGTATCGCGTAACGGC......GTGGCGGCATTGGCGGGCGAT 321
          :::|||:::|||   ||||||      ::::::   |||||||||:::
    150 ..GlyValIleSerValAsnGlyGlySerIleSerLeuLeuAlaGlyGln 165

    322 CAA.................................TATATTGTGAG 335
        :::                                    |||   :::::
    166 LysIleThrIleSerAspIleIleAsnProThrIleThrTyrSerIleAl 182

    336 CGTGGCACATAACGGCGGCTATAACAATGTTGATTTTGGTGCGGAGGGAA 385
        ::::   :::|||   :::   |||      |||:::   |||:::|||:
    182 aAlaProGluAsnGluAlaIleAsnLeuGlyAspIlePheAlaLysGlyG 199

    386 GCAATCCCGATCAGCACCGCTTTTCTTACCAAATTGTGAAAAGAAATAAT 435
        ::|||
    199 lyAsn.................................................. 200

    436 TATAAAGCAGGGACTAACGGCCATCCTTATGGCGGCGATTATCATATGCC 485

    200 .................................................... 200
```

```
 692 aGlyIleAsnMetAspSerIleAsnIleThrGlyGlyLeu..... 705

2084 CCCGCAATGTTGCCAAAGTGGAAGGCGATTGGCATTTAAGCAATCACGCC 2133
                              |||:::   :::::::::|||
 706 .........................AspPheSerIleThrSerHisAsn 713

2134 CAAGCAGTTTTCGGTGTCGCACCGCATCAAAGCCACACAATCTGTACACG 2183
     :::                      :::|||:::::::::      ::
 714 Arg.......................AsnSerAsnAlaPheGluIleLy 722

2184 TTCGGACTGGACGGGTCTGACAAGTTGTACCGAAAAAACCATT......A 2227
     :::|||   |||      ::::::   ::::::   ::::::      |
 722 sLysAspLeuThrIleAsnAlaThrGlySerAsnPheSerLeuLysGlnT 739

2228 CCGACGATAAAGTGATTGCTTCATTGAGCAAGACCGACATCAGAGGCAAT 2277
     ||   |||:::         :::   ||||||
 739 hrLysAspSerPheTyrAsnGluTyrSerLys.................. 749

2278 GTCAGCCTTGCCGATCACGCTCATTTAAATCTCACAGGACTTGCCACACT 2327
                    ||||||      :::      :::|||:::   ||
 750 ..............HisAlaIleAsnSerSerHisAsnLeuThrIleLe 761

2328 CAACGGCAATCTTAGTGCAGGCGGAGACACGCACTATACGGTTACGCGCA 2377
     |:::||||||::::::   ||||||:::                ::: : :
 761 uGlyGlyAsnValThrLeuGlyGlyGlu..............AsnSerS 773

2378 ACGCCACCCAAAACGGCAACCTCAGCCTCGTGGGCAATGCCCAAGCAACA 2427
     ::::::::   :::||||||:::::::::::::::::|||::::::|||
 773 erSerSerIleThrGlyAsnIleAsnIleThrAsnLysAlaAsnValThr 789

2428 TTTAATCAAGCCACATTAAACGGCAACACATCGGCTTCGGACAATGCTTC 2477
     ::::::      |||   |||:::||||||:::   :::   ::: . ::
 790 LeuGlnAlaAspThrSerAsnSerAsnThrGlyLeuLys...LysArgTh 805

2478 ATTTAATCTAAGCAACAACGCCGTACAAAACGGCAGTCTGACGCTTTCCG 2527
     :::::::|||:::|||   :::|||:::   |||:::|||:::|||:::
 805 rLeuThrLeuGlyAsnIleSerValGlu...GlyAsnLeuSerLeuThr. 820

2528 ACAACGCTAAGGCAAACGTAAGCCATTCCGCACTCAACGGCAATGTCTCC 2577
       :::|||:::|||||||:::                  ||||||:::|||
 821 ..GlyAlaAsnAlaAsnIleVal..............GlyAsnLeuSer 831

2578 CTAGCCGATAAGGCAGTATTCCATTTTGAA.........AACAGCCGCTT 2618
     :::|||:::   ::::::|||      |||        |||   :::::
 832 IleAlaGluAspSerThrPheLysGlyGluAlaSerAspAsnLeuAsnIl 848

2619 TACCGGAAAAATCAGCGGCGGCAAGGATACGGCATTACACTTAAAAGACA 2668
     :||||||   :::::::::
 848 eThrGlyThrPheThrAsn............................. 854

2669 GCGAATGGACGCTGCCGTCGGGCACGGAATTAGGCAATTTAAACCTTGAC 2718
                    :::|||||||      :::|||:::|||:::
 855 .................AsnGlyThr......AlaAsnIleAsnIleLys 863

2719 AACGCCACCATTACACTCAATTCCGCCTATCGACACGATGCGGCAGGCGC 2768
     ::::::::::::   |||::::::      ::::::   ::::::
 864 GlnGlyValValLysLeuGlnGlyAspIleAsnAsnLysGlyGly..... 878

2769 GCAAACCGGCAGTGCGGCAGATGCGCCGCGCCGCCGTTCGCGCCGTTCCC 2818

 878 ................................................. 878

2819 TATTATCCGTTACGCCGCCAACTTCGGCAGAATCCCGTTTCAACACGCTG 2868
       |||:::::|||      |||:::|||::::::      :::|||:::
 879 ..LeuAsnIleThr......ThrAsnAlaSerGlyThrGlnLysThrIle 892

2869 ACGGTAAACGGCAAATTG...AACGGTCAGGGAACATTCCGCTTTATGTC 2915
       :::||||||::::::   |||   :::|||   :::::::::   ::
 893 ...IleAsnGlyAsnIleThrAsnGluLysGlyAspLeuAsnIleLysAs 908

2916 G...................GAACTCTTCGGCTACCGCAGCGGCAAAT 2944
     :                    ::::::   |||      |||   |||
 908 nIleLysAlaAspAlaGluIleGlnIleGlyGlyAsnIleSerGlnLys. 924

2945 TGAAGCTGGCGGAAAGTTCCGAAGGCACTTACACCTTGGCT......... 2985
                           ||||||:::   |||::::::
 925 ....................GluGlyAsnLeuThrIleSerSerAspLys 934

2986 GTCAACAATACCGGCAACGAACCCGTAAGTCTCGAGCAATTGACGGTA.. 3033
     ||||||   |||                      :::|||:::|||:::
 935 ValAsnIleThr.....................AsnGlnIleThrIleLy 944

3034 .......GTGGAA...GGAAAAGACAACACACCGCTGTCCGAAAATCTTA 3073
             ||||||   |||:::::::::::      :::||||||   |
 944 sAlaGlyValGluGlyGlyArgSerAspSerSerGluAlaGluAsnAlaA 961

3074 ATTTCACCCTGCAAAACGAACACGTCGATGCCGGCGCATGGCGTTATCAG 3123
     ||:::|||:::|||
 961 snLeuThrIleGln................................... 965

3124 CTTATCCGCAAAGACGGCGAGTTCCGCCTGCATAATCCGGTCAAAGAACA 3173
                                                :::||||||
 966 ....................................ThrLysGluLe 969

3174 AGAGCTTTCCGACAAACTCGGCAAGGCGGGAGAAACAGAGGCCGCCTTGA 3223
      :::|||:::      |||:::   :::|||   :::::|||   :::|
 969 uLysLeuAlaGlyAspLeuAsnIleSerGlyPheAsnLysAlaGluIleT 986

3224 CGGCAAAACAGGCACAACTTGCCGCCAAACAACAGGCGGAAAAAGACAAC 3273
     ||||||||::::::::      :::      :::|||:::      |||
 986 hrAlaLysAsnGlySerAspLeuThrIleGlyAsnAlaSerGlyGlyAsn 1002

3274 GCGCAAAGC.......................CTTGACGCGCTGATTGC 3299
     |||::::::                       |||:::  |||::
1003 AlaAspAlaLysLysValThrPheAspLysValLysAspSerLysIleSe 1019

3300 GGCC...GGGCGCAATGCCACCGAAAAGGCAGAA...AGTGTTGCCGAAC 3343
     ::::   |||:::|||:::|||   ::::::|||   ::::::::::::
1019 rThrAspGlyHisAsnValThrLeuAsnSerGluValLysThrSerAsnG 1036

3344 CGGCCCGGCAGGCAGGCGGGGAAAATGCC...GGCATTATGCAGGCGGAG 3390
       :::   :::|||||:::::::|||:::   |||:::      :::
1036 lySerSerAsnAlaGlyAsnAspAsnSerThrGlyLeuThrIleSerAla 1052

3391 GAAGAG..........................AAAAAACGGGTGCAGGC 3413
     ::::::                         |||   ::::::
1053 LysAspValThrValAsnAsnAsnValThrSerHisLysThrIleAsnIl 1069

3414 GGATAAAGACACCGCCTTGGCGAAACAGCGCGAAGCGGAAACCCGGCCGG 3463
      :::      ::::::   :::      :::|||:::   |||      |
1069 eSerAlaAlaAlaGlyAsnValThrThrLysGluGlyThrThrIleAsnA 1086

3464 CTACCACCGCCTTCCCCCGCGCCCGCCGCGCCCGCCGGGATTTGCCGCAA 3513
     ||||||||:::  |||          ::|||       |||      :::::
1086 laThrThrGly.SerValGluValThrAlaGlnAsnGlyThrIleLysGl 1102

3514 CCGCAGCCCCAACCGCAACCCCAACCGCAGCGCGACCTGATCAGCCGTTA 3563
     ::::      ::::::|||   :::   :::||||||   :::   |||
1102 yAsnIleThrSerGlnAsnValThrValThrAlaThrGluAsnLeuValT 1119

3564 TGCCAATAGCGGTTTGAGTGAATTTTCCGCCACGCTCAACAGCGTTTTCG 3613
                  :::   :::|||           ::::::   :::|||
1119 hrThrGluAsnAlaValIleAsnAlaThrSerGlyThrValAsnIleSer 1135

3614 CCGTACAGGACGAATTGGACCGCGTGTTTGCCGAAGACCGCCGCAACGCC 3663
                                       ||||||:::
1136 ............................ThrLysThrGlyAspIleLy 1142

3664 GTTTGGA...CAAGCGGCATCCGGGACACCAAACACTACCGTTCGC.... 3706
        |||   :::::::::||||||:::   |||   |||:::
1142 sGlyGlyIleGluSerThrSerGlyAsnValAsnIleThrAlaSerGlyA 1159

3707 ........AAGATTTCCGCGCCTACCGCCAACAAACCGACCTGCGCCAAA 3748
             |||:::|||      |||::::::      |||   :::
1159 snThrLeuLysValSerAsnIleThrGlyGlnAspValThrValThrAla 1175
```

```
    3749 TCGGTATGCAGAAAAACCTCGGCAGCGGGCGCGTCGGCATCCTGTTTTCG 3798
         ::::::         |||::::::||||||:::::   |||:::
    1176 AspAlaGlyAlaLeuThrThrThrAlaGlySerThrIleSerAla..... 1190

    3799 CACAACCGGA..................................... 3808
          |||||||||
    1191 .ThrThrGlyAsnAlaAsnIleThrThrLysThrGlyAspIleAsnGlyL 1207

    3809 ........CCGGAAACACCTTCGACGACGGCATCGGCAACTCGGCACGGC 3850
                  ::::::   |||:::|||      ||||||
    1207 ysValGluSerSerSerGlySerValThrLeuValAlaThrGlyAlaThr 1223

    3851 TTGCCCACGGTGCCGTTTTCGGGCAAT............ACGGCATCGGC 3888
         |||   :::      :::||||||||||           ||||||:::::
    1224 LeuAlaValGlyAsnIleSerGlyAsnThrValThrIleThrAlaAspSe 1240

    3889 AGGTTCGACATCGGCATCAGCGCGGGCGCGGGTTTTAGTAGCGGCAGCCT 3938
         :|||:::   ::::::::::::   :::   :::      :::   :::
    1240 rGlyLysLeuThrSerThrValGlySerThrIleAsnGlyThrAsnServ 1257

    3939 TTCAGACGGCATCAGAGGCAAAA.................TCCGCCGCC 3970
              |||:::|||::::::                     |||:::
    1257 alThrThrSerSerGlnSerGlyAspIleGluGlyThrIleSerGlyAsn 1273

    3971 GCGTGCTGCATTACGGCATTCAGGCAAGAT..............ACCGC 4005
         :::      :::||||||         |||               :::||
    1274 ThrValAsnValThrAlaSerThrGlyAspLeuThrIleGlyAsnSerAl 1290

    4006 GCAGGTTTCGGCGGATTCGGCATCGAACCGCACATCGGCGCAACGCGCTA 4055
         |:::|||:::|||   :::::::::         :::|||:::
    1290 aLysValGluAlaLysAsnGlyAlaAlaThrLeuThrAlaGlu....... 1304

    4056 TTTCGTCCAAAAAGCGGATTACCGATACGAAAACGTCAATATCGCCACCC 4105
              |||   |||      |||:::|||   :::||||||:::
    1305 .....SerGlyLysLeuThrThrGlnThrGlySerSerIleThrSerSer 1319

    4106 CGGGCCTTGCATTCAACCGCTACCGCGCGG 4135
         ::::::   ::::::|||   |||||||:::
    1320 AsnGly...GlnThrThrLeuThrAlaLys 1328

seq_name: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-08-460-269C-6

seq_documentation_block:
; Sequence 6, Application US/08460269C
; Patent No. 6197548
;   GENERAL INFORMATION:
;        APPLICANT: CLARE, JEFFREY J.
;                   ROMANOS, MICHAEL A.
;        TITLE OF INVENTION: EXPRESSION OF HETEROLOGOUS PROTEIN IN
;                            YEAST
;        NUMBER OF SEQUENCES: 17
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Millen, White, Zelano & Branigan, P.C.
;             STREET: 2200 Clarendon Blvd., Suite 1400
;             CITY: ARLINGTON
;             STATE: VA
;             COUNTRY: USA
;             ZIP: 22201
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/08/460,269C
;             FILING DATE: 02-Jun-1995
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Lebovitz, Richard M.
;             REGISTRATION NUMBER: 37,067
;             REFERENCE/DOCKET NUMBER: Popov-2
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (703) 243-6333
;             TELEFAX: (703) 243-6410

;   INFORMATION FOR SEQ ID NO: 6:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 922 amino acids
;             TYPE: amino acid
;             TOPOLOGY: linear
;        MOLECULE TYPE: protein
;        SEQUENCE DESCRIPTION: SEQ ID NO: 6:
US-08-460-269C-6


alignment_scores:
           Quality:  275.50            Length:   1162
             Ratio:  0.566               Gaps:     54
 Percent Similarity:  41.910   Percent Identity: 21.687

alignment_block:
 US-09-303-518D-653 x US-08-460-269C-6   ..

 Align seg 1/1  to: US-08-460-269C-6  from: 1  to: 922

    1366 AACGACCGCCTGTCCAAAATCGGCAAAGGC.............ACGCT 1400
         |||      ||||||:::|||   |||:::             |||
       2 AsnMetSerLeuSerArgIleValLysAlaAlaProLeuArgArgThrTh 18

    1401 GCTGGTTCAAGCCAAAGGG.............................. 1419
          |||:::::|||   |||
      18 rLeuAlaMetAlaLeuGlyAlaLeuGlyAlaAlaProAlaAlaTyrAlaA 35

    1420 .....GAAAACCAAGGCTCGGTCAGCGTGGGCGACGGTAAA...GTCATC 1461
               :::||||||:::   :::::::::|||:::   :::      |||
      35 spTrpAsnAsnGlnSerIleIleLysAlaGlyGluArgGlnHisGlyIle 51

    1462 TTAGATCAGCAGGCGGACGATCAAGGCAAAAAACAAGCCTTTAGTGAAAT 1511
              :::|||:::|||      |||   :::   |||   :::
      52 HisIleLysGlnSerAspGlyAlaGlyValArgThrAlaThrGlyThrTh 68

    1512 CGGCTTGGTCAGCGGCAGGGGGACGGTGCAACTGAATGCCGATAATCAGT 1561
              ||||||||||||   :::       :::       :::|||
      68 rIleLysValSerGlyArgGlnAlaGlnGlyValLeuLeuGluAsn.... 83

    1562 TCAACCCCGACAAACTCTATTTCGGCTTTCGCGGCGGACGTTTGGATTTG 1611
                              :::   |||::::::|||   :::
      84 ........ProAlaAlaGluLeuArgPheGlnAsnGlySerValThrSer 97

    1612 AACGGGCATTCGCTTTCGTTCCACCGCATTCAAAATACCGATGAAGGGGC 1661
         :::|||                        |||      |||||||||::
      98 SerGly.......................GlnLeuPheAspGluGlyVa 106

    1662 GATGATTGTCAACCACAATCAAGACAAAGAATCCACCGTTACCATTACAG 1711
         :                    :::         :::||||||||||:::   :
     106 l.................ArgArgPheLeuGlyThrValThrValLysA 117

    1712 GCAATAAA.........GATATTACTACAACCGGCAATAACAACAACTTG 1752
         :::::|||         |||   :::|||   :::|||   ::::::
     117 laGlyLysLeuValAlaAspHisAlaThrLeuAlaAsnValSerAspThr 133

    1753 GATAGCAAAAAAGAAATTGCC...TACAACGGTTGGTTTGGCGAGAAAGA 1799
           :::         ||||||   |||   :::       ||||||:::
     134 ArgAspAspAspGlyIleAlaLeuTyrValAla......GlyGluGln.. 147

    1800 TGCAACCAAAACGAACGGGCGGCTCAATCTGAATTACCAACCGGAAGAAG 1849
                                                |||   :::   |
     148 ..............................AlaGlnAlaSerIleA 153

    1850 CGGATCGCACTTTACTGCTTTCCGGCGGAACAAATTTA......AACGGC 1893
         |||||   ||||||      :::||||||:::::::::      ::::::
     153 laAspSerThrLeuGlnGlyAlaGlyGlyValArgValGluArgGlyAla 169

    1894 AATATCACGCAAACAAACGGCAAACTGTTTTTCAGCGGCAGACCGACACC 1943
         |||:::|||      ::::::   :::      :::|||
     170 AsnValThrValGlnArgSerThrIleValAspGlyGlyLeu........ 183

    1944 GCACGCCTACAATCATTTAGGAAGCGGGTGGTCA...AAAATGGAAGGTA 1990
```

```
                       |||:::|||:::                 :::    |||    :
 184 .............HisIleGlyThrLeuGlnProLeuGlnProGluAspL 196

1991 TCCCACAAGGAGAAATCGTGTGGGACAACGATTGGATCGACCGC...... 2034
      ::|||   :::::::::|||        :::      :::
 196 euProProSerArgValValLeuGlyAspThrSerValThrAlaValPro 212

2035 ..............................ACATTTAAAGCGGAAAACTTCCA 2057
                                   :::|||   |||:::::::::
 213 AlaSerGlyAlaProAlaAlaValPheValPheGlyAlaAsnGluLeuTh 229

2058 TATTCAGGGCGGACAAGCGGTGGTTTCCCGCAATGTTGCCAAAGTGGAAG 2107
      ::::::||||||
 229 rValAspGlyGly..................................... 233

2108 GCGATTGGCATTTAAGCAATCACGCCCAAGCAGTTTTCGGTGTCGCACCG 2157
              |||:::::::::          |||        |||||||||
 234 ........HisIleThrGlyGlyArgAlaAla......GlyValAla... 244

2158 CATCAAAGCCACACAATCTGTACACGTTCGGACTGGACGGGTCTGACAAG 2207

 244 .................................................. 244

2208 TTGTACCGAAAAAACCATTACCGACGATAAAGTGATTGCTTCATTGAGCA 2257
                           :::   |||   :::|||:::   |||:::
 245 ..................AlaMetAspGlyAlaIleValHisLeuGlnA 255

2258 AGACCGACATCAGAGGCAATGTCAGCCTTGCCGATCACGCTCATTTAAAT 2307
      :::::  ||||||
 255 rgAlaThrIleArg.................................... 259

2308 CTCACAGGACTTGCCACACTCAACGGCAATCTTAGTGCAGGCGGAGACAC 2357
                               :::|||:::    |||||||||
 260 .....................ArgGlyAspAlaProAlaGlyGly..... 267

2358 GCACTATACGGTTACGCGCAACGCCACCCAAAACGGCAACCTC....... 2400
              :::|||      :::|||:::   :::|||   :::
 268 .......AlaValProGlyGlyAlaValProGlyGlyAlaValProGlyG 282

2401 .............................................AGC 2403
                                                  |||
 282 lyPheGlyProLeuLeuAspGlyTrpTyrGlyValAspValSerAspSer 298

2404 CTCGTGGGCAATGCCCAAGCAACATTTAATCAAGCCACATTAAACGGCAA 2453
        |||      ||||||:::        :::         |||::::::
 299 ThrValAspLeuAlaGlnSerIleValGluAlaProGlnLeuGlyAlaAl 315

2454 CACATCGGCTTCGGACAATGCTTCATTTAATCTAAGC............. 2490
             |||:::   :::|||        :::::::|||
 315 aIleArgAlaGlyArgGlyAlaArgValThrValSerGlyGlySerLeuS 332

2491 ..........AACAACGCCGTACAAAACGGC.................. 2511
              :::|||::::::::::::|||
 332 erAlaProHisGlyAsnValIleGluThrGlyGlyGlyAlaArgArgPhe 348

2512 .....................AGTCTGACGCTTTCCGACAACGCTAAGGC 2540
                          |||:::||||||:::   :::|||:::||
 349 ProProProAlaSerProLeuSerIleThrLeuGlnAlaGlyAlaArgAl 365

2541 AAAC..................................GTAAGCCATT 2554
      |:::                                  |||:::   :
 365 aGlnGlyArgAlaLeuLeuTyrArgValLeuProGluProValLysLeuT 382

2555 CCGCACTCAACGGCAATGTCTCCCTAGCCGATAAGGCAGTATTCCATTTT 2604
      ::       :::|||      :::   :::|||      :::
 382 hrLeuAlaGlyGlyAlaGlnGlyGlnGlyAsp......IleValAlaThr 396

2605 GAAAACAGCCGCTTTACCGGAAAAATCAGCGGCGGCAAGGATACGGCATT 2654
      |||         :::   |||      ||||||      |||:::|||||
 397 GluLeuProProIleProGlyAlaSerSerGlyProLeuAspValAlaLe 413

2655 ACACTTAAAAGACAGCGAATGGACGCTGCCGTCGGGCACGGAATTAGGCA 2704
      |        :::::::::::||||||        :::        :::   :

 413 u...AlaSerGlnAlaArgTrpThr...GlyAlaThrArgAlaValAspS 428

2705 ATTTAAACCTTGACAACGCCACCATTACACTCAATTCCGCCTATCGACAC 2754
      ::|||:::::::||||||||||||     :::::::::::
 428 erLeuSerIleAspAsnAlaThrTrpValMetThrAsp........... 440

2755 GATGCGGCAGGCGCGCAAACCGGCAGTGCGGCAGATGCGCCGCGCCGCCG 2804
                               :::|||   :::    |||    |||
 441 .....................AsnSerAsnValGlyAlaLeuArgLeuAl 450

2805 TTCGCGCCGTTCCCTATTATCCGTTACGCCGCCAACTTCGGCAGAATCCC 2854
       |||      |||    ::::::            |||::::::|||    :::|
 450 aSerAspGlySer...ValAspPheGlnGlnProAlaGluAla...GlyA 465

2855 GTTTCAACACGCTGACGGTAAACGGCAAATTGAACGGTCAGGGAACATTC 2904
      |||||:::::|||      |||:::          |||    |||:::|||    |||
 465 rgPheLysValLeuMetValAsp...ThrLeuAlaGlySerGlyLeuPhe 480

2905 CGCTTTATGTCGGAACTCTTCGGCTACCGCAGCGGCAAATTGAAGCTGGC 2954
      |||:::                            |||    |||||    :::
 481 ArgMetAsnValPheAlaAspLeuGlyLeuSerAspLysLeuValValMe 497

2955 GGAAAGTTCCGAAGGCACTTACACCTTGGCTGTCAACAATACCGGCAACG 3004
       ::::::::::::|||    :::    |||    |||:::|||:::|||:::|
 497 tArgAspAlaSerGlyGlnHisArgLeuTrpValArgAsnSerGlySerG 514

3005 AACCCGTAAGTCTCGAGCAATTGACGGTAGTGGAAGGAAAAGACAACACA 3054
      |||||:::|||    :::    :::    :::|||:::             |||
 514 luProAlaSerGlyAsnThrMetLeuLeuValGln...........Thr 526

3055 CCGCTGTCCGAAAATCTTAATTTCACCCTGCAAAACGAA......CACGT 3098
      |||    ::::::         :::||||||||||   |||:::              ||
 527 ProArgGlySerAlaAlaThrPheThrLeuAlaAsnLysAspGlyLysVa 543

3099 CGATGCCGGCGCATGGCGTTATCAGCTTATCCGCAAA...GACGGCGAGT 3145
      ||||    |||::::::|||||||:::|||          :::    :::|||::::
 543 lAspIleGlyThrTyrArgTyrArgLeuAlaAlaAsnGlyAsnGlyGlnT 560

3146 TCCGCCTGCATAATCCGGTCAAAGAACAAGAGCTTTCCGACAAACTCGGC 3195
      ::    |||  :::
 560 rpSerLeu.Val...................................... 563

3196 AAGGCGGGAGAAACAGAGGCCGCCTTGACGGCAAAACAGGCACAACTTGC 3245

 563 .................................................. 563

3246 CGCCAAACAACAGGCGGAAAAAGACAACGCGCAAAGCCTTGACGCGCTGA 3295

 563 .................................................. 563

3296 TTGCGGCCGGGCGCAATGCCACCGAAAAGGCAGAAAGTGTTGCCGAACCG 3345
                   ||||||         |||          |||         |||   ::
 564 .........GlyAlaLysAlaProProAlaProLysProAlaPro...Gl 576

3346 GCCCGGCAGGCAGGCGGGGAAAATGCCGGCATTATGCAGGCGGAGGAAGA 3395
      :||||||    |||    |||    :::|||
 576 nProGlyProGlnProGlyProGlnProPro.................. 586

3396 GAAAAAACGGGTGCAGGCGGATAAAGACACCGCCTTGGCGAAACAGCGCG 3445

 586 .................................................. 586

3446 AAGCGGAAACCCGGCCGGCTACCACCGCCTTCCCCCGCGCCCGCCGCGCC 3495
                           ||||||      ||||||:::||||||:::|
 587 ..................GlnProProGln.ProProGlnProProGlnP 597

3496 CGCCGGGATTTGCCGCAACCGCAGCCCCAA.........CCGCAACCCCA 3536
      ||||| :::    ||||||    ||||||:::             |||||||||
 597 roPro.GlnProProGlnArgGlnProGluAlaProAlaProGlnProPr 613

3537 ACCGCAGCGCGACCTGATCAGCCGTTATGCCAATAGCGGTTTGAGTGAAT 3586
              |||:::|||     |||          ||||||:::::::::::
 613 oAlaGlyArgGluLeu...SerAlaAlaAlaAsnAlaAlaValAsnThrG 629
```

```
3587 TTTCCGCCACGCTCAACAGCGTT......TTCGCCGTACAGGACGAATTG 3630
      ::::::   |||   |||:::      :::|||   ::::::   |||
 629 lyGlyValGlyLeuAlaSerThrLeuTrpTyrAlaGluSerAsnAlaLeu 645

3631 GACCGCGTGTTTGCCGAAGACCGCCGCAAC............GCCGTTTG 3668
     ::::::   ::::::|||   |||   |||          ::::::||
 646 SerLysArgLeuGlyGluLeuArgLeuAsnProAspAlaGlyGlyAlaTr 662

3669 GACAAGCGGCATCCGGGACACCAAACACTACCGTTCGCAAGATTTCCGCG 3718
     |      |||:::                                    |
 662 pGlyArgGlyPhe.................................A 667

3719 CCTACCGCCAACAAACCGACCTGCGCCAAATCGGTATGCAGAAAAACCTC 3768
     ||   |||||||||   |||
 667 laGlnArgGlnGlnLeuAsp.......................... 673

3769 GGCAGCGGGCGCGTCGGCATCCTGTTTTCGCACAACCGGACCGGAAACAC 3818
                                 ||||||:::|||:::
 674 ..............................AsnArgAlaGlyArgAr 679

3819 CTTCGACGAC...................GGCATCGGCAACTCGGCAC 3847
      |||||||:::                |||      :::::::::
 679 gPheAspGlnLysValAlaGlyPheGluLeuGlyAlaAspHisAlaValA 696

3848 GGCTTGCCCACGGTGCC......GTTTTCGGGCAATACGGCATCGGCAGG 3891
       :::|||   |||         :::   |||      |||      |||
 696 laValAlaGlyGlyArgTrpHisLeuGlyGlyLeuAlaGlyTyrThrArg 712

3892 TTCGACATCGGCATCAGCGCGGGCGCGGGTTTTAGTAGCGGCAGCCTT.. 3939
        |||   |||:::::::::   :::|||      :::   |||:::
 713 GlyAspArgGlyPheThrGlyAspGlyGlyGlyHisThrAspSerValHi 729

3939 ..................................................... 3939

 729 sValGlyGlyTyrAlaThrTyrIleAlaAsnSerGlyPheTyrLeuAspA 746

3940 ..............................................TCA 3942
                                                    |||
 746 laThrLeuArgAlaSerArgLeuGluAsnAspPheLysValAlaGlySer 762

3943 GACGGC......ATCAGAGGCAAAATCCGCCGCCGCGTGCTGCATTACGG 3986
     ||||||      ::::::||||||   |||   :::   :::      ::
 763 AspGlyTyrAlaValLysGlyLysTyrArgThrHisGlyValGlyValSe 779

3987 CATTCAGGCAAGATACCGC...GCAGGTTTCGGCGGATTCGGCATCGAAC 4033
     ::::::::|||      |||   |||         |||:::   :::||||
 779 rLeuGluAlaGlyArgArgPheAlaHisAlaAspGlyTrpPheLeuGluP 796

4034 CGCACATCGGCGCAACGCGCTATTTCGTCCAAAAAGCGGATTACCGATAC 4083
     ||:::         :::   :::   |||      :::   ||||||
 796 roGlnAlaGluLeuAlaValPheArgValGlyGlyGlyAlaTyrArgAla 812

4084 GAAAAC...GTCAATATCGCCACCCCGGGCCTTGCATTCAACCGCTACCG 4130
        |||   :::::::::         |||   :::          ||
 813 AlaAsnGlyLeuArgValArgAspGluGlyGlySerSerValLeuGlyAr 829

4131 CGCGGGCATTAAGGCAGATTATTCATTCAAACCGGCGCAACACATTTCCA 4180
     |   |||:::::::::         ::::::   |||          ::::
 829 gLeuGlyLeuGluValGlyLysArgIleGluLeuAlaGlyGlyArgGlnV 846

4181 TCACGCCTTATTTGAGCCTGTCCTATACCGATGCCGCTTCCGGCAAAGTC 4230
     ::   ||||||:::::   |||      :::      :::|||
 846 alGlnProTyrIleLysAlaSerValLeuGlnGluPheAspGlyAlaGly 862

4231 CGAACGCGCGTCAATACCGCCGTATTGGCGCAGGATTTCGGCAAAACCCG 4280
        :::|||:::|||     :::       ::::::     |||||
 863 ThrValArgThrAsnGlyIleAlaHisArgThrGluLeuArgGlyThrAr 879

4281 CAGTGCGGAATGGGGCGTAAACGCCGAAATCAAAGGTTTCACGCTGTCCC 4330
     |   ||||||   |||                          |||::::
 879 g...AlaGluLeuGly.........................LeuGlyM 886
```

```
4331 TCCACGCTGCCGCCGCCAAGGGGCCGCAATTGGAAGCG......CAGCAC 4374
     ::   ||||||   :::::::|||   :::|||   |||      ::::::
 886 etAlaAlaAlaLeuGlyArgGlyHisSerLeuTyrAlaSerTyrGluTyr 902

4375 AGCGCGGGCATCAAATTAGGCTACCGCTGG 4404
     |||   |||   ||||||:::      |||
 903 SerLysGlyProLysLeuAlaMetProTrp 912
```

```
seq_name: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:US-08-038-682-4

seq_documentation_block:
; Sequence 4, Application US/08038682
; Patent No. 5549897
;  GENERAL INFORMATION:
;    APPLICANT:  BARENKAMP, STEPHEN J
;    APPLICANT:  ST. GEME III, JOSEPH W
;    TITLE OF INVENTION:  HIGH MOLECULAR WEIGHT SURFACE PROTEINS
;    TITLE OF INVENTION:  OF NON-TYPEABLE HAEMOPHILUS
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Shoemaker and Mattare, Ltd
;      STREET:  2001 Jefferson Davis Hwy., 1203 Crystal Plaza
;      STREET:  Bldg. 1
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22202-0286
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/038,682
;      FILING DATE:  16-MAR-1993
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BERKSTRESSER, JERRY W
;      REGISTRATION NUMBER:  22,651
;      REFERENCE/DOCKET NUMBER:  1038-293
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 415-0810
;      TELEFAX:  (703) 415-0813
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1477 amino acids
;      TYPE:  amino acids
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-038-682-4


alignment_scores:
            Quality:  274.50              Length:   1488
              Ratio:  0.419                 Gaps:     73
 Percent Similarity:  44.019   Percent Identity: 19.220

alignment_block:
 US-09-303-518D-653 x US-08-038-682-4   ..

 Align seq 1/1  to: US-08-038-682-4  from: 1  to: 1477

     178 AATAAAGGCAAGTTTGCAGTCGGGGCGAAAGATATTGAGGTTTACAACAA 227
         |||   |||   ::::::||||||   ||||||   :::|||
     199 AsnHisGlyLeuIleThrValGly...LysAspGlySerVal........ 211

     228 AAAAGGGGAGTTGGTCGGCAAATCGATGACGAAAGCCCCGATGATTGATT 277
                 :::|||:::|||   ::::::   :::
     212 .......AsnLeuIleGlyGlyLysValLysAsnGlu............. 221

     278 TTTCTGTGGTATCGCGTAACGGC......GTGGCGGCATTGGCGGGCGAT 321
           :::|||:::|||   ||||||      ::::::   |||||||||:::
```

```
222 ..GlyValIleSerValAsnGlyGlySerIleSerLeuLeuAlaGlyGln 237

322 CAA.....TATATTGTGAG 335
238 LysIleThrIleSerAspIleIleAsnProThrIleThrTyrSerIleAl 254

336 CGTGGCACATAACGGCGGCTATAACAATGTTGATTTTGGTGCGGAGGGAA 385
254 aAlaProGluAsnGluAlaValAsnLeuGlyAspIlePheAlaLysGlyG 271

386 GCAAT.....CCCGATCAGCACCGCTTTTCT 411
271 lyAsnIleAsnValArgAlaAlaThrIleArgAsnGlnGlyLysLeuSer 287

412 TACCAAATTGTGAAAAGAAATAATTATAAAGCAGGGACTAACGGCCATCC 461
288 AlaAspSerValSerLysAsp......LysSerGly..... 297

462 TTATGGCGGCGATTATCATATGCCGCGTTTGCACAAATTTGTCACAGATG 511
298 .....AsnIleValLeuSerA 303

512 CAGAACCTGTTGAG.....ATGACCAGTTATATGGAT 543
303 laLysGluGlyGluAlaGluIleGlyGlyValIleSerAlaGlnAsnGln 319

544 GGGTGGAAATACGCTGATTTAAATAAATACCCTGATCGTGTT.....CG 587
320 GlnAlaLysGlyGlyLysLeuMetIleThrGlyAspLysValThrLeuLy 336

588 AATCGGAGCA.....GGCAGACAATATT 610
336 sThrGlyAlaValIleAspLeuSerGlyLysGluGlyGlyGluThrTyrL 353

611 GGCGGTCTGATGAA...GACGAACCCAATAAC..... 639
353 euGlyGlyAspGluArgGlyGluGlyLysAsnGlyIleGlnLeuAlaLys 369

640 .....CGCGAAAGTTCATATCATATTGCAAGC..... 666
370 LysThrSerLeuGluLysGlySerThrIleAsnValSerGlyLysGluLy 386

667 .....GCATATTCTTGG.....CTCGTCGGTGGCAATA 694
386 sGlyGlyPheAlaIleValTrpGlyAspIleAlaLeuIleAspGlyAsnI 403

695 CCTTTGCACAAAATGGATCAGGTGGTGGCACAGTCAACTTAGGTAGCGAA 744
403 leAsnAlaGln...GlySerGly..... 409

745 AAAATTAAACATAGCCCATATGGTTTTTTACCAACAGGAGGCTCATTTGG 794
410 .....AspIleAlaLysThrGlyGlyPheValGl 419

795 CGACAGTGGCTCACCAATGTTTATCTATGATGCCCAAAAGCAAAAGTGGT 844
419 uThrSerGlyHisAspLeuPheIleLysAsp..... 429

845 TAATTAATGGGGTATTGCAAACAGGCAACCCCTATATAGGAAAAAGCAAT 894
430 .....AsnAlaIleValAspAla..... 435

895 GGCTTCCAGCTAGTTCGTAAAGATTGGTTCTATGATGAAATCTTTGCTGG 944
436 .....LysGluTrpLeuLeuAsp.....Ph 442

945 AGATACCCATTCAGTATTCTACGAA...CCACATCAAAATGGGAAATACT 991
442 eAspAsnValSerIleAsnAlaGluAspProLeuPheAsnAsnThrGlyI 459

992 TTTTTAACGACAATAATAATGGCGCAGGAAAAATCGATGCCAAACATAAA 1041
459 leAsnAspGluPheProThrGlyThrGlyGluAlaSerAspProLysLys 475

1042 CACTATTCTCTACCTTATAGATTAAAAACACGAACCGTTCAATTGTTT.. 1089
476 AsnSerGluLeuLysThrThrLeuThrAsnThrThrIleSerAsnTyrLe 492

1090 .....AATGTTTCTTTATCCGAGACAGCAAGAGAAC 1120
492 uLysAsnAlaTrpThrMetAsnIleThrAlaSerArgLysLeuThr.... 507

1121 CTGTTTATCATGCTGCAGGTGGGGTCAACAGTTATCGACCCAGACTGAAT 1170
508 .....ValAsnSer.....SerIleAsn 513

1171 AATGGAGAAAATATTTCCTTTATTGACAAAGGAAAAGGTGAATTGATACT 1220
514 IleGlySerAsnSerHisLeuIleLeuHisSerLysGly..... 526

1221 TACCAGCAACATCAACCAAGGCGCGGGCGGTTTGTATTTTGAGGGTAATT 1270
527 .....GlnArgGlyGlyGlyValGlnIleAspGlyAspI 538

1271 TTACGGTCTCGCCTAAAAACAACGAAACGTGGCAAGGCGCGGGCGTTCAT 1320
538 leThrSerLysGlyGlyAsnLeuThrIleTyrSerGlyGlyTrpValAsp 554

1321 ATCAGTGATGGCAGTACCGTTACTTGGAAAGTAAACGGCGTGGCAAACGA 1370
555 ValHisLysAsnIleThr..... 560

1371 CCGCCTGTCCAAAATCGGCAAAGGCACGCTGCTGGTTCAAGCC..... 1413
561 .....LeuAspGlnGlyPheLeuAsnIleThrAlaAlaServ 573

1414 .....AAAGGGGAAAACCAAGGCTCGGTCAGCGTGGGCGACGGTAAA 1455
573 alAlaPheGluGlyGlyAsnAsnLysAlaArgAspAlaAlaAsnAlaLys 589

1456 GTCATCTTAGATCAGCAGGCGGACGATCAAGGCAAAAAACAAGCCTTTAG 1505
590 IleVal..... 591

1506 TGAAATCGGCTTGGTCAGCGGCAGGGGGACGGTGCAACTGAATGCCGATA 1555
592 .....AlaGlnGlyThrValThrIleThrGlyGluG 602

1556 ATCAGTTCAACCCCGACAAACTCTATTTCGGCTTTCGCGGCGGACGTTTG 1605
602 lyLys.....AspPheArgAlaAsnAsnVal 610

1606 GATTTGAACGGGCATTCGCTTTCGTTCCACCGCATTCAAAATACCGATGA 1655
611 SerLeuAsnGly.....ThrGlyLy 617

1656 AGGGGCGATGATTGTC.....AACCACAATCAAGACAAAGAAT 1693
617 sGlyLeuAsnIleIleSerSerValAsnAsnLeuThrHisAsnLeuSerG 634

1694 CCACCGTTACCATTACAGGCAATAAAGATATTACTACAACCGGCAATAAC 1743
634 lyThrIleAsnIleSerGlyAsnIleThrIleAsnGlnThrThrArgLys 650

1744 AAC..... 1746
651 AsnThrSerTyrTrpGlnThrSerHisAspSerHisTrpAsnValSerAl 667

1747 ....AACTTGGAT..... 1755
667 aLeuAsnLeuGluThrGlyAlaAsnPheThrPheIleLysTyrIleSerS 684

1756 .....AGCAAAAAAGAAATTGCC 1773
684 erAsnSerLysGlyLeuThrThrGlnTyrArgSerSerAlaGlyValAsn 700
```

```
   1116 SerAspAsnAspThrGlyLeuThrIleThrAlaLysAsnValGluVal.. 1131

   3453 AACCCGGCCGGCTACCACCGCCTTCCCCCGCGCCCGCCGCGCCCGCCGGG 3502

   1131 ..................................................  1131

   3503 ATTTGCCGCAACCGCAGCCCCAACCGCAACCCCAACCGCAGCGCGACCTG 3552
                                            ::::::|||:::
   1132 .....................................AsnLysAspIle 1135

   3553 ATCAGC...CGTTATGCCAATAGCGGTTTGAGTGAATTTTCCGCCACGCT 3599
           |||   :::   :::|||         ||||||      :::||| :
   1136 ThrSerLeuLysThrValAsnIleThrAlaSerGluLysValThrThr.T 1152

   3600 CAACAGCGTTTTCGCCGTACAGGACGAATTGGACCGCGTGTTTGCCGAAG 3649
        ::::::::   :::         ||||||      |||   :::
   1152 hrAlaGlySerThrIleAsnAlaThrAsnGlyLysAlaSerIleThrThr 1168

   3650 ACCGCCGCAACGCCGTTTGGACAAGCGGCATCCGGGACACCAAACACTAC 3699
           ::::::         |||      :::::::::|||   ::::::::
   1169 LysThrGlyAspIleSerGlyThrIleSerGlyAsnThrValSerValSe 1185

   3700 CGTTCGC...AAGATTTCCGCGCCTACCGCCAACAAACCGACCTGCGCCA 3746
        ::::         :::::::::   :::::::::|||      :::   :
   1185 rAlaThrValAspLeuThrThrLysSerGlySerLysIleGluAlaLysS 1202

   3747 AATCGGTATGCAGAAAAACCTCGGCAGCGGGCGCGTCGGCATCCTGTTTT 3796
        :::::   ::::::   |||||||||:::|||:::   ::::::
   1202 erGlyGluAlaAsnValThrSerAlaThrGlyThrIleGlyGly...... 1216

   3797 CGCACAACCGGACCGGAAACACCTTCGACGACGGCATCGGCA........ 3838
           |||   :::   :::|||   ::::::|||||||:::|||
   1217 ...ThrIleSerGlyAsnThrValAsnValThrAlaAsnAlaGlyAspLe 1232

   3839 ..........................ACTCGGCACGGCTTGCCCACGG 3860
                                       :::   :::|||      |||:
   1232 uThrValGlyAsnGlyAlaGluIleAsnAlaThrGluGlyAlaAlaThrL 1249

   3861 TGCCGTTTTCGGGCAAT......ACGGCATCGGCAGGTTCGACATCGGCA 3904
        ::      :::||||||      |||::::::||||||||||:::   :::
   1249 euThrAlaThrGlyAsnThrLeuThrThrGluAlaGlySerSerIleThr 1265

   3905 TCAGCGCGGGCGCGGGTTTTAGTAGCGGCAGCCTTTCAGACGGCATCAGA 3954
        |||:::::::::::::|||   :::   |||      |||:::::|||
   1266 SerThrLysGlyGlnValAspLeuLeuAla......GlnAsnGlySerIl 1280

   3955 GGCAAAATCCGCCGCCGCGTGCTGCATTACGGCATTCAGGCAAGATACCG 4004
         |||   |||      |||:::   :::|||               |||
   1280 eAlaGlySerIleAsnAlaAlaAsnValThrLeuAsnThrThrGlyThrL 1297

   4005 CGCAGGTTTCGGCGGATTCGGCATCGAACCGCACATCGGCGCAACGCGCT 4054
           :::   |||   |||      :::   ||||||:::      :::
   1297 euThrThrValAlaGlySerAspIleLysAlaThrSerGlyThrLeuVal 1313

   4055 ATTTCGTCCAAA 4066
        |||:::::::|||
   1314 IleAsnAlaLys 1317

seq_name: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:US-08-302-832-4

seq_documentation_block:
; Sequence 4, Application US/08302832
; Patent No. 5603938
;  GENERAL INFORMATION:
;    APPLICANT:  Barenkamp, Stephen J
;    TITLE OF INVENTION:  High Molecular Weight Surface Proteins
;    TITLE OF INVENTION:  of No. 5603938-Typeable Haemophilus
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Shoemaker and Mattare, Ltd.
;      STREET:  2001 Jefferson Davis Hwy., 1203 Crystal Plaza
;      STREET:  Bldg. 1
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22202-0286
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/302,832
;      FILING DATE:  16-SEP-1994
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9205704.1
;      FILING DATE:  16-MAR-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US pct/us93/02166
;      FILING DATE:  16-MAR-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Berkstresser, Jerry W
;      REGISTRATION NUMBER:  22,651
;      REFERENCE/DOCKET NUMBER:  1038-404
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 415-0810
;      TELEFAX:  (703) 415-0813
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1477 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-302-832-4


alignment_scores:
            Quality:  274.50              Length:   1488
              Ratio:  0.419                 Gaps:     73
 Percent Similarity:  44.019   Percent Identity: 19.220

alignment_block:
 US-09-303-518D-653 x US-08-302-832-4   ..

 Align seg 1/1  to: US-08-302-832-4  from: 1  to: 1477

     178 AATAAAGGCAAGTTTGCAGTCGGGGCGAAAGATATTGAGGTTTACAACAA 227
         |||   |||   :::::|||||   ||||||   :::|||
     199 AsnHisGlyLeuIleThrValGly...LysAspGlyServal........ 211

     228 AAAAGGGGAGTTGGTCGGCAAATCGATGACGAAAGCCCCGATGATTGATT 277
                 :::|||:::|||   ::::::   :::
     212 .......AsnLeuIleGlyGlyLysValLysAsnGlu............. 221

     278 TTTCTGTGGTATCGCGTAACGGC......GTGGCGGCATTGGCGGGCGAT 321
           :::|||:::|||   ||||||      ::::::   ||||||||||:::
     222 ..GlyValIleSerValAsnGlyGlySerIleSerLeuLeuAlaGlyGln 237

     322 CAA.................................TATATTGTGAG 335
         :::                                    |||   :::::
     238 LysIleThrIleSerAspIleIleAsnProThrIleThrTyrSerIleAl 254

     336 CGTGGCACATAACGGCGGCTATAACAATGTTGATTTTGGTGCGGAGGGAA 385
         ::::   :::|||   :::   |||      |||:::   |||:::|||:
     254 aAlaProGluAsnGluAlaValAsnLeuGlyAspIlePheAlaLysGlyG 271

     386 GCAAT......................CCCGATCAGCACCGCTTTTCT 411
         ::|||                          :::|||   ::::::|||
     271 lyAsnIleAsnValArgAlaAlaThrIleArgAsnGlnGlyLysLeuSer 287

     412 TACCAAATTGTGAAAAGAAATAATTATAAAGCAGGGACTAACGGCCATCC 461
         :::   |||:::::::::      |||:::|||
     288 AlaAspSerValSerLysAsp......LysSerGly.............. 297

     462 TTATGGCGGCGATTATCATATGCCGCGTTTGCACAAATTTGTCACAGATG 511
```


298 ..... AsnIleValLeuSerA 303
512 CAGAACCTGTTGAG..... ATGACCAGTTATATGGAT 543
303 laLysGluGlyGluAlaGluIleGlyGlyValIleSerAlaGlnAsnGln 319
544 GGGTGGAAATACGCTGATTTAAATAAATACCCTGATCGTGTT.....CG 587
320 GlnAlaLysGlyGlyLysLeuMetIleThrGlyAspLysValThrLeuLy 336
588 AATCGGAGCA..... GGCAGACAATATT 610
336 sThrGlyAlaValIleAspLeuSerGlyLysGluGlyGlyGluThrTyrL 353
611 GGCGGTCTGATGAA...GACGAACCCAATAAC..... 639
353 euGlyGlyAspGluArgGlyGluGlyLysAsnGlyIleGlnLeuAlaLys 369
640 .....CGCGAAAGTTCATATCATATTGCAAGC..... 666
370 LysThrSerLeuGluLysGlySerThrIleAsnValSerGlyLysGluLy 386
667 .....GCATATTCTTGG..... CTCGTCGGTGGCAATA 694
386 sGlyGlyPheAlaIleValTrpGlyAspIleAlaLeuIleAspGlyAsnI 403
695 CCTTTGCACAAAATGGATCAGGTGGTGGCACAGTCAACTTAGGTAGCGAA 744
403 leAsnAlaGln...GlySerGly..... 409
745 AAAATTAAACATAGCCCATATGGTTTTTTACCAACAGGAGGCTCATTTGG 794
410 ..... AspIleAlaLysThrGlyGlyPheValGl 419
795 CGACAGTGGCTCACCAATGTTTATCTATGATGCCCAAAAGCAAAAGTGGT 844
419 uThrSerGlyHisAspLeuPheIleLysAsp..... 429
845 TAATTAATGGGGTATTGCAAACAGGCAACCCCTATATAGGAAAAAGCAAT 894
430 .....AsnAlaIleValAspAla..... 435
895 GGCTTCCAGCTAGTTCGTAAAGATTGGTTCTATGATGAAATCTTTGCTGG 944
436 ..... LysGluTrpLeuLeuAsp.....Ph 442
945 AGATACCCATTCAGTATTCTACGAA...CCACATCAAAATGGGAAATACT 991
442 eAspAsnValSerIleAsnAlaGluAspProLeuPheAsnAsnThrGlyI 459
992 TTTTTAACGACAATAATAATGGCGCAGGAAAAATCGATGCCAAACATAAA 1041
459 leAsnAspGluPheProThrGlyThrGlyGluAlaSerAspProLysLys 475
1042 CACTATTCTCTACCTTATAGATTAAAAACACGAACCGTTCAATTGTTT.. 1089
476 AsnSerGluLeuLysThrThrLeuThrAsnThrThrIleSerAsnTyrLe 492
1090 ..... AATGTTTCTTTATCCGAGACAGCAAGAGAAC 1120
492 uLysAsnAlaTrpThrMetAsnIleThrAlaSerArgLysLeuThr.... 507
1121 CTGTTTATCATGCTGCAGGTGGGGTCAACAGTTATCGACCCAGACTGAAT 1170
508 ..... ValAsnSer.....SerIleAsn 513
1171 AATGGAGAAAATATTTCCTTTATTGACAAAGGAAAAGGTGAATTGATACT 1220
514 IleGlySerAsnSerHisLeuIleLeuHisSerLysGly..... 526
1221 TACCAGCAACATCAACCAAGGCGCGGGCGGTTTGTATTTTGAGGGTAATT 1270
527 ..... GlnArgGlyGlyGlyValGlnIleAspGlyAspI 538
1271 TTACGGTCTCGCCTAAAAACAACGAAACGTGGCAAGGCGCGGGCGTTCAT 1320
538 leThrSerLysGlyGlyAsnLeuThrIleTyrSerGlyGlyTrpValAsp 554
1321 ATCAGTGATGGCAGTACCGTTACTTGGAAAGTAAACGGCGTGGCAAACGA 1370
555 ValHisLysAsnIleThr..... 560
1371 CCGCCTGTCCAAAATCGGCAAAGGCACGCTGCTGGTTCAAGCC..... 1413
561 ..... LeuAspGlnGlyPheLeuAsnIleThrAlaAlaServ 573
1414 .....AAAGGGGAAAACCAAGGCTCGGTCAGCGTGGGCGACGGTAAA 1455
573 alAlaPheGluGlyGlyAsnAsnLysAlaArgAspAlaAlaAsnAlaLys 589
1456 GTCATCTTAGATCAGCAGGCGGACGATCAAGGCAAAAAACAAGCCTTTAG 1505
590 IleVal..... 591
1506 TGAAATCGGCTTGGTCAGCGGCAGGGGGACGGTGCAACTGAATGCCGATA 1555
592 ..... AlaGlnGlyThrValThrIleThrGlyGluG 602
1556 ATCAGTTCAACCCCGACAAACTCTATTTCGGCTTTCGCGGCGGACGTTTG 1605
602 lyLys.....AspPheArgAlaAsnAsnVal 610
1606 GATTTGAACGGGCATTCGCTTTCGTTCCACCGCATTCAAAATACCGATGA 1655
611 SerLeuAsnGly..... ThrGlyLy 617
1656 AGGGGCGATGATTGTC..... AACCACAATCAAGACAAAGAAT 1693
617 sGlyLeuAsnIleIleSerServalAsnAsnLeuThrHisAsnLeuSerG 634
1694 CCACCGTTACCATTACAGGCAATAAAGATATTACTACAACCGGCAATAAC 1743
634 lyThrIleAsnIleSerGlyAsnIleThrIleAsnGlnThrThrArgLys 650
1744 AAC..... 1746
651 AsnThrSerTyrTrpGlnThrSerHisAspSerHisTrpAsnValSerAl 667
1747 ....AACTTGGAT..... 1755
667 aLeuAsnLeuGluThrGlyAlaAsnPheThrPheIleLysTyrIleSerS 684
1756 ..... AGCAAAAAAGAAATTGCC 1773
684 erAsnSerLysGlyLeuThrThrGlnTyrArgSerSerAlaGlyValAsn 700
1774 TACAACGGTTGGTTTGGCGAGAAAGATGCAACCAAAACGAACGGGCGGCT 1823
701 PheAsnGly..... ValAsnGlyAsnMe 708
1824 CAATCTGAATTACCAACCGGAAGAAGCGGATCGCACTTTACTGCTTTCCG 1873
708 tSerPheAsnLeuLys...GluGlyAlaLysValAsnPheLysLeuLysP 724
1874 GCGGAACAAATTTAAAC..... GGC 1893
724 roAsnGluAsnMetAsnThrSerLysProLeuProIleArgPheLeuAla 740
1894 AATATCACGCAAACA...AACGGCAAACTGTTTTTCAGCGGCAGACCGAC 1940
741 AsnIleThrAlaThrGlyGlyGlySerValPhePheAsp..... 753
1941 ACCGCACGCCTAC...AATCATTTAGGAAGCGGGTGGTCAAAAATGGAAG 1987
754 .....IleTyrAlaAsnHisSerGlyArgGly..... AlaG 764


1988 GTATCCCACAAGGAGAAATCGTGTGGGACAACGATTGGATCGACCGCACA 2037
764 luLeuLysMetSerGluIleAsnIleSerAsn 774
2038 TTTAAAGCGGAAAACTTC CATATTCAGGGCGGACAAGC 2075
775 GlyAlaAsnPheThrLeuAsnSerHisValArgGlyAspAspAl 789
2076 GGTGGTTTCCCGCAATGTTGCCAAAGTGGAAGGCGATTGGCATTTAAGCA 2125
789 aPhe LysIleAsnLysAspLeuThrIleAsn. 799
2126 ATCACGCCCAAGCAGTTTTCGGTGTCGCACCGCATCAAAGCCACACAATC 2175
799 799
2176 TGTACACGTTCGGACTGGACGGGTCTGACAAGTTGTACCGAAAAAACCAT 2225
800 AlaThrAsnSerAsnPhe 805
2226 TACCGACGATAAAGTGATTGCTTCATTGAGCAAGACCGACATCAGAGGCA 2275
806 SerLeuArgGlnThrLysAspAspPheT 815
2276 ATGTCAGCCTTGCCGATCACGCTCATTTAAATCTCACAGGACTTGCCACA 2325
815 yrAspGlyTyrAlaArgAsnAlaIleAsnSerThrTyrAsnIleSerIle 831
2326 CTCAACGGCAATCTTAGTGCAGGCGGAGACACGCACTATACGGTTACGCG 2375
832 LeuGlyGlyAsnValThrLeuGlyGly 840
2376 CAACGCCACCCAAAACGGCAACCTCAGCCTCGTGGGCAATGCCCAAGCAA 2425
841 GlnAsnSerSerSerSerIleThrGlyAsn IleT 852
2426 CATTTAATCAAGCC ACATTAAACGGCAACACATCGGCTTCG 2466
852 hrIleGluLysAlaAlaAsnValThrLeuGluAlaAsnAsnAlaProAsn 868
2467 GACAATGCTTCATTTAATCTAAGCAACAACGCCGTACAAAACGGCAGTCT 2516
869 GlnGln AsnIleArgAspArgValIleLysLeuGlySerLe 882
2517 GACGCTTTCCGACAACGCTAAGGCAAACGTAAGCCATTCCGCACTCAACG 2566
882 uLeuValAsnGlySerLeuSerLeuThrGlyGluAsnAlaAspIleLysG 899
2567 GCAATGTCTCCCTAGCCGATAAGGCAGTATTCCAT TTTGAA 2607
899 lyAsnLeuThrIleSerGluSerAlaThrPheLysGlyLysThrArgAsp 915
2608 AACAGCCGCTTTACCGGAAAAATCAGCGGCGGCAAGGATACGGCATTACA 2657
916 ThrLeuAsnIleThrGlyAsnPheThrAsnAsnGlyThrAlaGluIleAs 932
2658 CTTAAAAGACAGCGAATGGACGCTGCCGTCGGGCACGGAATTAGGCAATT 2707
932 nIleThrGlnGlyValValLysLeuGlyAsnValThrAsnAspGlyAspL 949
2708 TAAACCTTGACAACGCCACCATTACACTCAATTCCGCCTATCGACACGAT 2757
949 euAsn IleThrThrHisAla 955
2758 GCGGCAGGCGCGCAAACCGGCAGTGCGGCAGATGCGCCGCGCCGCCGTTC 2807
956 LysArgAs 958
2808 GCGCCGTTCCCTATTATCCGTTACGCCGCCAACTTCGGCAGAATCCCGTT 2857
958 nGlnArgSerIle 962
2858 TCAACACGCTGACGGTAAACGGCAAATTGAACGGTCAGGGAACATTCCGC 2907
963 IleGlyGlyAspIleIleAsn 969
2908 TTTATGTCGGAACTCTTCGGCTACCGCAGCGGCAAATTGAAGCTGGCGGA 2957
970 LysLysGlySerLeuAsnIleThrAs 978
2958 AAGTTCC G 2965
978 pSerAsnAsnAspAlaGluIleGlnIleGlyGlyAsnIleSerGlnLysG 995
2966 AAGGCACTTACACCTTGGCT GTCAACAATACCGGCAACGAA 3006
995 luGlyAsnLeuThrIleSerSerAspLysIleAsnIleThr 1008
3007 CCCGTAAGTCTCGAGCAATTGACGGTA GTGGAAGGAAAAGA 3047
1009 LysGlnIleThrIleLysLysGlyIleAspGlyGluAs 1021
3048 C AACACACCGCTGTCCGAAAATCTTAATTTCACCCTGCAAAACGAAC 3094
1021 pSerSerSerAspAlaThrSerAsnAlaAsnLeuThrIleLys 1035
3095 ACGTCGATGCCGGCGCATGGCGTTATCAGCTTATCCGCAAAGACGGCGAG 3144
1035 1035
3145 TTCCGCCTGCATAATCCGGTCAAAGAACAAGAGCTTTCCGACAAACTCGG 3194
1036 ThrLysGluLeuLysLeuThrGluAspLeuSe 1046
3195 CAAGGCGGGAGAAACAGAGGCCGCCTTGACGGCAAAACAGGCACAACTTG 3244
1046 rIleSerGlyPheAsnLysAlaGluIleThrAlaLysAspGlyArgAspL 1063
3245 CCGCCAAACAACAGGCGGAAAAAGACAACGCGCAAAGC 3282
1063 euThrIleGlyAsnSerAsnAspGlyAsnSerGlyAlaGluAlaLysThr 1079
3283 CTTGACGCGCTGATTGCGGCC GGGCGCAA 3311
1080 ValThrPheAsnAsnValLysAspSerLysIleSerAlaAspGlyHisAs 1096
3312 TGCCACCGAAAAGGCAGAAAGTGTTGCCGAACCGGCCCGGCAGGCAGGCG 3361
1096 nValThr LeuAsnSerLysValLysThrSerSerSerAsnGlyG 1111
3362 GGGAAAATGCCGGCATTATGCAGGCGGAGGAAGAGAAAAAACGGGTGCAG 3411
1111 lyArgGluSer Asn 1115
3412 GCGGATAAAGACACCGCCTTG GCGAAACAGCGCGAAGCGGA 3452
1116 SerAspAsnAspThrGlyLeuThrIleThrAlaLysAsnValGluVal.. 1131
3453 AACCCGGCCGGCTACCACCGCCTTCCCCCGCGCCCGCCGCGCCCGCCGGG 3502
1131 1131
3503 ATTTGCCGCAACCGCAGCCCCAACCGCAACCCCAACCGCAGCGCGACCTG 3552
1132 AsnLysAspIle 1135
3553 ATCAGC CGTTATGCCAATAGCGGTTTGAGTGAATTTTCCGCCACGCT 3599
1136 ThrSerLeuLysThrValAsnIleThrAlaSerGluLysValThrThr.T 1152
3600 CAACAGCGTTTTCGCCGTACAGGACGAATTGGACCGCGTGTTTGCCGAAG 3649
1152 hrAlaGlySerThrIleAsnAlaThrAsnGlyLysAlaSerIleThrThr 1168
3650 ACCGCCGCAACGCCGTTTGGACAAGCGGCATCCGGGACACCAAACACTAC 3699

```
        ::::::          |||       :::::::::|||   ::::::::
    1169 LysThrGlyAspIleSerGlyThrIleSerGlyAsnThrValSerValSe 1185

    3700 CGTTCGC...AAGATTTCCGCGCCTACCGCCAACAAACCGACCTGCGCCA 3746
         ::::         :::::::::   :::::::::|||      :::   :
    1185 rAlaThrValAspLeuThrThrLysSerGlySerLysIleGluAlaLysS 1202

    3747 AATCGGTATGCAGAAAAACCTCGGCAGCGGGCGCGTCGGCATCCTGTTTT 3796
         :::::   ::::::   |||||||||:::|||:::   ::::::
    1202 erGlyGluAlaAsnValThrSerAlaThrGlyThrIleGlyGly..... 1216

    3797 CGCACAACCGGACCGGAAACACCTTCGACGACGGCATCGGCA........ 3838
            |||   :::   :::|||   ::::::|||||:::|||
    1217 ...ThrIleSerGlyAsnThrValAsnValThrAlaAsnAlaGlyAspLe 1232

    3839 ..........................ACTCGGCACGGCTTGCCCACGG 3860
                                     :::   :::|||      |||:
    1232 uThrValGlyAsnGlyAlaGluIleAsnAlaThrGluGlyAlaAlaThrL 1249

    3861 TGCCGTTTTCGGGCAAT......ACGGCATCGGCAGGTTCGACATCGGCA 3904
         ::      :::||||||      |||:::::|||||||||:::   :::
    1249 euThrAlaThrGlyAsnThrLeuThrThrGluAlaGlySerSerIleThr 1265

    3905 TCAGCGCGGGCGCGGGTTTTAGTAGCGGCAGCCTTTCAGACGGCATCAGA 3954
         |||:::::::::::::|||   :::   |||      |||:::::|||
    1266 SerThrLysGlyGlnValAspLeuLeuAla......GlnAsnGlySerIl 1280

    3955 GGCAAAATCCGCCGCCGCGTGCTGCATTACGGCATTCAGGCAAGATACCG 4004
          |||   |||      |||:::   :::|||                |||
    1280 eAlaGlySerIleAsnAlaAlaAsnValThrLeuAsnThrThrGlyThrL 1297

    4005 CGCAGGTTTCGGCGGATTCGGCATCGAACCGCACATCGGCGCAACGCGCT 4054
              :::   |||   |||      :::   ||||||:::      :::
    1297 euThrThrValAlaGlySerAspIleLysAlaThrSerGlyThrLeuVal 1313

    4055 ATTTCGTCCAAA 4066
         |||::::::|||
    1314 IleAsnAlaLys 1317

seq_name: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:US-08-530-198-4

seq_documentation_block:
; Sequence 4, Application US/08530198
; Patent No. 5869065
;  GENERAL INFORMATION:
;    APPLICANT:  BARENKAMP, STEPHEN J
;    APPLICANT:  ST. GEME III, JOSEPH W
;    TITLE OF INVENTION:  HIGH MOLECULAR WEIGHT SURFACE PROTEINS
;    TITLE OF INVENTION:  OF NON-TYPEABLE HAEMOPHILUS
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Shoemaker and Mattare, Ltd
;      STREET:  2001 Jefferson Davis Hwy., 1203 Crystal Plaza
;      STREET:  Bldg. 1
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22202-0286
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/530,198
;      FILING DATE:  13-DEC-1995
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BERKSTRESSER, JERRY W
;      REGISTRATION NUMBER:  22,651
;      REFERENCE/DOCKET NUMBER:  JWB-1186
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 415-0810
;      TELEFAX:  (703) 415-0813
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1477 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-530-198-4


alignment_scores:
            Quality:  274.50              Length:   1488
              Ratio:  0.419                 Gaps:     73
 Percent Similarity:  44.019   Percent Identity: 19.220

alignment_block:
 US-09-303-518D-653 x US-08-530-198-4   ..

 Align seg 1/1  to: US-08-530-198-4  from: 1  to: 1477

     178 AATAAAGGCAAGTTTGCAGTCGGGGCGAAAGATATTGAGGTTTACAACAA 227
         |||   |||   ::::::||||||   ||||||   :::|||
     199 AsnHisGlyLeuIleThrValGly...LysAspGlySerVal........ 211

     228 AAAAGGGGAGTTGGTCGGCAAATCGATGACGAAAGCCCCGATGATTGATT 277
               :::|||:::|||   ::::::   :::
     212 .......AsnLeuIleGlyGlyLysValLysAsnGlu.............. 221

     278 TTTCTGTGGTATCGCGTAACGGC......GTGGCGGCATTGGCGGGCGAT 321
           :::|||:::|||   ||||||      ::::::   |||||||||:::
     222 ..GlyValIleSerValAsnGlyGlySerIleSerLeuLeuAlaGlyGln 237

     322 CAA.................................TATATTGTGAG 335
         :::                                 |||   :::::
     238 LysIleThrIleSerAspIleIleAsnProThrIleThrTyrSerIleAl 254

     336 CGTGGCACATAACGGCGGCTATAACAATGTTGATTTTGGTGCGGAGGGAA 385
         ::::   :::|||   :::   |||      |||:::   |||:::|||:
     254 aAlaProGluAsnGluAlaValAsnLeuGlyAspIlePheAlaLysGlyG 271

     386 GCAAT......................CCCGATCAGCACCGCTTTTCT 411
         ::|||                       :::|||   ::::::|||
     271 lyAsnIleAsnValArgAlaAlaThrIleArgAsnGlnGlyLysLeuSer 287

     412 TACCAAATTGTGAAAAGAAATAATTATAAAGCAGGGACTAACGGCCATCC 461
            :::   |||:::::::::      |||:::|||
     288 AlaAspSerValSerLysAsp......LysSerGly.............. 297

     462 TTATGGCGGCGATTATCATATGCCGCGTTTGCACAAATTTGTCACAGATG 511
                                       ::::::|||   :::|
     298 ..............................AsnIleValLeuSerA 303

     512 CAGAACCTGTTGAG.................ATGACCAGTTATATGGAT 543
         ||:::     |||                 :::::::::      :::
     303 laLysGluGlyGluAlaGluIleGlyGlyValIleSerAlaGlnAsnGln 319

     544 GGGTGGAAATACGCTGATTTAAATAAATACCCTGATCGTGTT......CG 587
               |||   :::   |||            |||:::|||     ::
     320 GlnAlaLysGlyGlyLysLeuMetIleThrGlyAspLysValThrLeuLy 336

     588 AATCGGAGCA.........................GGCAGACAATATT 610
         :   ||||||                        |||:::   |||
     336 sThrGlyAlaValIleAspLeuSerGlyLysGluGlyGlyGluThrTyrL 353

     611 GGCGGTCTGATGAA...GACGAACCCAATAAC...:.............. 639
               :::||||||      |||   :::|||
     353 euGlyGlyAspGluArgGlyGluGlyLysAsnGlyIleGlnLeuAlaLys 369

     640 ..............CGCGAAAGTTCATATCATATTGCAAGC........ 666
                        :::   |||:::   ::::::::::
     370 LysThrSerLeuGluLysGlySerThrIleAsnValSerGlyLysGluLy 386
```

```
667 .........GCATATTCTTGG...........CTCGTCGGTGGCAATA 694
            |||     |||           |||:::  ||||||
386 sGlyGlyPheAlaIleValTrpGlyAspIleAlaLeuIleAspGlyAsnI 403

695 CCTTTGCACAAAATGGATCAGGTGGTGGCACAGTCAACTTAGGTAGCGAA 744
      ||||||  |||||||||
403 leAsnAlaGln...GlySerGly..................... 409

745 AAAATTAAACATAGCCCATATGGTTTTTTACCAACAGGAGGCTCATTTGG 794
                          :::     |||||||||
410 ..................AspIleAlaLysThrGlyGlyPheValGl 419

795 CGACAGTGGCTCACCAATGTTTATCTATGATGCCCAAAAGCAAAAGTGGT 844
      ||||||     :::|||||||  |||
419 uThrSerGlyHisAspLeuPheIleLysAsp.................. 429

845 TAATTAATGGGGTATTGCAAACAGGCAACCCCTATATAGGAAAAAGCAAT 894
      |||:::::::::::::::
430 .....AsnAlaIleValAspAla...................... 435

895 GGCTTCCAGCTAGTTCGTAAAGATTGGTTCTATGATGAAATCTTTGCTGG 944
                    |||:::|||:::  |||
436 ..................LysGluTrpLeuLeuAsp...........Ph 442

945 AGATACCCATTCAGTATTCTACGAA...CCACATCAAAATGGGAAATACT 991
  |||:::  |||:::     |||  |||     |||:::     :
442 eAspAsnValSerIleAsnAlaGluAspProLeuPheAsnAsnThrGlyI 459

992 TTTTTAACGACAATAATAATGGCGCAGGAAAAATCGATGCCAAACATAAA 1041
    ::  ::::::     :::|||:::|||:::  :::         |||
459 leAsnAspGluPheProThrGlyThrGlyGluAlaSerAspProLysLys 475

1042 CACTATTCTCTACCTTATAGATTAAAAACACGAACCGTTCAATTGTTT.. 1089
     :::  :::|||        |||  :::  |||::::::  :::
476 AsnSerGluLeuLysThrThrLeuThrAsnThrThrIleSerAsnTyrLe 492

1090 .................AATGTTTCTTTATCCGAGACAGCAAGAGAAC 1120
                     |||::::::  |||:::
492 uLysAsnAlaTrpThrMetAsnIleThrAlaSerArgLysLeuThr.... 507

1121 CTGTTTATCATGCTGCAGGTGGGGTCAACAGTTATCGACCCAGACTGAAT 1170
                          |||||||||          :::|||
508 ....................ValAsnSer.........SerIleAsn 513

1171 AATGGAGAAAATATTTCCTTTATTGACAAAGGAAAAGGTGAATTGATACT 1220
     |||:::|||     :::|||     :::||||||
514 IleGlySerAsnSerHisLeuIleLeuHisSerLysGly........... 526

1221 TACCAGCAACATCAACCAAGGCGCGGGCGGTTTGTATTTTGAGGGTAATT 1270
                   |||  :::||||||:::  ::::::|||:::
527 ..............GlnArgGlyGlyGlyValGlnIleAspGlyAspI 538

1271 TTACGGTCTCGCCTAAAAACAACGAAACGTGGCAAGGCGCGGGCGTTCAT 1320
     ::|||  :::      |||         ::::::|||:::  |||
538 leThrSerLysGlyGlyAsnLeuThrIleTyrSerGlyGlyTrpValAsp 554

1321 ATCAGTGATGGCAGTACCGTTACTTGGAAAGTAAACGGCGTGGCAAACGA 1370
     :::     :::  |||
555 ValHisLysAsnIleThr............................... 560

1371 CCGCCTGTCCAAAATCGGCAAAGGCACGCTGCTGGTTCAAGCC..... 1413
                 :::  :::|||  |||  :::  |||
561 ............LeuAspGlnGlyPheLeuAsnIleThrAlaAlaServ 573

1414 ........AAAGGGGAAAACCAAGGCTCGGTCAGCGTGGGCGACGGTAAA 1455
           :::|||  |||:::  :::  ::::::::::::::|||
573 alAlaPheGluGlyGlyAsnAsnLysAlaArgAspAlaAlaAsnAlaLys 589

1456 GTCATCTTAGATCAGCAGGCGGACGATCAAGGCAAAAAACAAGCCTTTAG 1505
     ::::::
590 IleVal.......................................... 591

1506 TGAAATCGGCTTGGTCAGCGGCAGGGGGACGGTGCAACTGAATGCCGATA 1555
```

```
                     ::::::|||||||||  ::::::::::::
592 ..................AlaGlnGlyThrValThrIleThrGlyGluG 602

1556 ATCAGTTCAACCCCGACAAACTCTATTTCGGCTTTCGCGGCGGACGTTTG 1605
     :::::                  ||||||::::::::::::
602 lyLys.....................AspPheArgAlaAsnAsnVal 610

1606 GATTTGAACGGGCATTCGCTTTCGTTCCACCGCATTCAAAATACCGATGA 1655
     :::||||||||||                          |||  ::
611 SerLeuAsnGly.............................ThrGlyLy 617

1656 AGGGGCGATGATTGTC............AACCACAATCAAGACAAAGAAT 1693
     :|||      |||:::            |||  ::::::::  ::::
617 sGlyLeuAsnIleIleSerSerValAsnAsnLeuThrHisAsnLeuSerG 634

1694 CCACCGTTACCATTACAGGCAATAAAGATATTACTACAACCGGCAATAAC 1743
     ::|||:::::|||:::|||||      |||:::  |||  ::::::
634 lyThrIleAsnIleSerGlyAsnIleThrIleAsnGlnThrThrArgLys 650

1744 AAC........................................... 1746
     |||
651 AsnThrSerTyrTrpGlnThrSerHisAspSerHisTrpAsnValSerAl 667

1747 ....AACTTGGAT................................. 1755
         ||||||:::
667 aLeuAsnLeuGluThrGlyAlaAsnPheThrPheIleLysTyrIleSerS 684

1756 ............................AGCAAAAAAGAAATTGCC 1773
                                |||:::      :::
684 erAsnSerLysGlyLeuThrThrGlnTyrArgSerSerAlaGlyValAsn 700

1774 TACAACGGTTGGTTTGGCGAGAAAGATGCAACCAAAACGAACGGGCGGCT 1823
     :::||||||                    :::||||||:::::
701 PheAsnGly.......................ValAsnGlyAsnMe 708

1824 CAATCTGAATTACCAACCGGAAGAAGCGGATCGCACTTTACTGCTTTCCG 1873
     :::::::|||   :::   |||  |||      ::::::   |||:::
708 tSerPheAsnLeuLys...GluGlyAlaLysValAsnPheLysLeuLysP 724

1874 GCGGAACAAATTTAAAC...........................GGC 1893
       :::   |||:::|||                           :::
724 roAsnGluAsnMetAsnThrSerLysProLeuProIleArgPheLeuAla 740

1894 AATATCACGCAAACA...AACGGCAAACTGTTTTTCAGCGGCAGACCGAC 1940
     |||||||||   |||   :::|||::::::|||||:::
741 AsnIleThrAlaThrGlyGlyGlySerValPhePheAsp........... 753

1941 ACCGCACGCCTAC...AATCATTTAGGAAGCGGGTGGTCAAAAATGGAAG 1987
               |||   ||||||  |||  |||
754 .......IleTyrAlaAsnHisSerGlyArgGly...........AlaG 764

1988 GTATCCCACAAGGAGAAATCGTGTGGGACAACGATTGGATCGACCGCACA 2037
       :::   ::::::||||||      :::|||
764 luLeuLysMetSerGluIleAsnIleSerAsn................. 774

2038 TTTAAAGCGGAAAACTTC............CATATTCAGGGCGGACAAGC 2075
           :::   ||||||           |||:::::::|||   :::||
775 ......GlyAlaAsnPheThrLeuAsnSerHisValArgGlyAspAspAl 789

2076 GGTGGTTTCCCGCAATGTTGCCAAAGTGGAAGGCGATTGGCATTTAAGCA 2125
     |                    |||::::::   |||      ::::::
789 aPhe..................LysIleAsnLysAspLeuThrIleAsn. 799

2126 ATCACGCCCAAGCAGTTTTCGGTGTCGCACCGCATCAAAGCCACACAATC 2175

799 ............................................... 799

2176 TGTACACGTTCGGACTGGACGGGTCTGACAAGTTGTACCGAAAAAACCAT 2225
     :::|||:::|||::::::
800 AlaThrAsnSerAsnPhe.............................. 805

2226 TACCGACGATAAAGTGATTGCTTCATTGAGCAAGACCGACATCAGAGGCA 2275
                          ||||||   :::|||
```

```
    3905 TCAGCGCGGGCGCGGGTTTTAGTAGCGGCAGCCTTTCAGACGGCATCAGA 3954
         |||:::::::::::|||   :::   |||      |||:::::::|||
    1266 SerThrLysGlyGlnValAspLeuLeuAla......GlnAsnGlySerIl 1280

    3955 GGCAAAATCCGCCGCCGCGTGCTGCATTACGGCATTCAGGCAAGATACCG 4004
          |||   |||      |||:::   :::|||               |||
    1280 eAlaGlySerIleAsnAlaAlaAsnValThrLeuAsnThrThrGlyThrL 1297

    4005 CGCAGGTTTCGGCGGATTCGGCATCGAACCGCACATCGGCGCAACGCGCT 4054
              :::   |||   |||      :::   ||||||:::      :::
    1297 euThrThrValAlaGlySerAspIleLysAlaThrSerGlyThrLeuVal 1313

    4055 ATTTCGTCCAAA 4066
         |||:::::::|||
    1314 IleAsnAlaLys 1317

seq_name: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:US-08-469-880-4

seq_documentation_block:
; Sequence 4, Application US/08469880
; Patent No. 5876733
;  GENERAL INFORMATION:
;    APPLICANT:  Barenkamp, Stephen J.
;    TITLE OF INVENTION:  High Molecular Weight Surface Proteins
;    TITLE OF INVENTION:  of No. 5876733-Typeable Haemophilus
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Shoemaker and Mattare, Ltd.
;      STREET:  2001 Jefferson Davis Hwy., 1203 Crystal Plaza
;      STREET:  Bldg. 1
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22202-0286
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/469,880
;      FILING DATE:  06-JUN-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9205704.1
;      FILING DATE:  16-MAR-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US PCT/US93/02166
;      FILING DATE:  16-MAR-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/302,832
;      FILING DATE:  16-SEP-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Berkstresser, Jerry W
;      REGISTRATION NUMBER:  22,651
;      REFERENCE/DOCKET NUMBER:  1038-516 MIS:vg
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 415-0810
;      TELEFAX:  (703) 415-0813
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1477 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-469-880-4


alignment_scores:
            Quality:  274.50            Length:   1488
              Ratio:  0.419               Gaps:     73
 Percent Similarity:  44.019   Percent Identity: 19.220
```

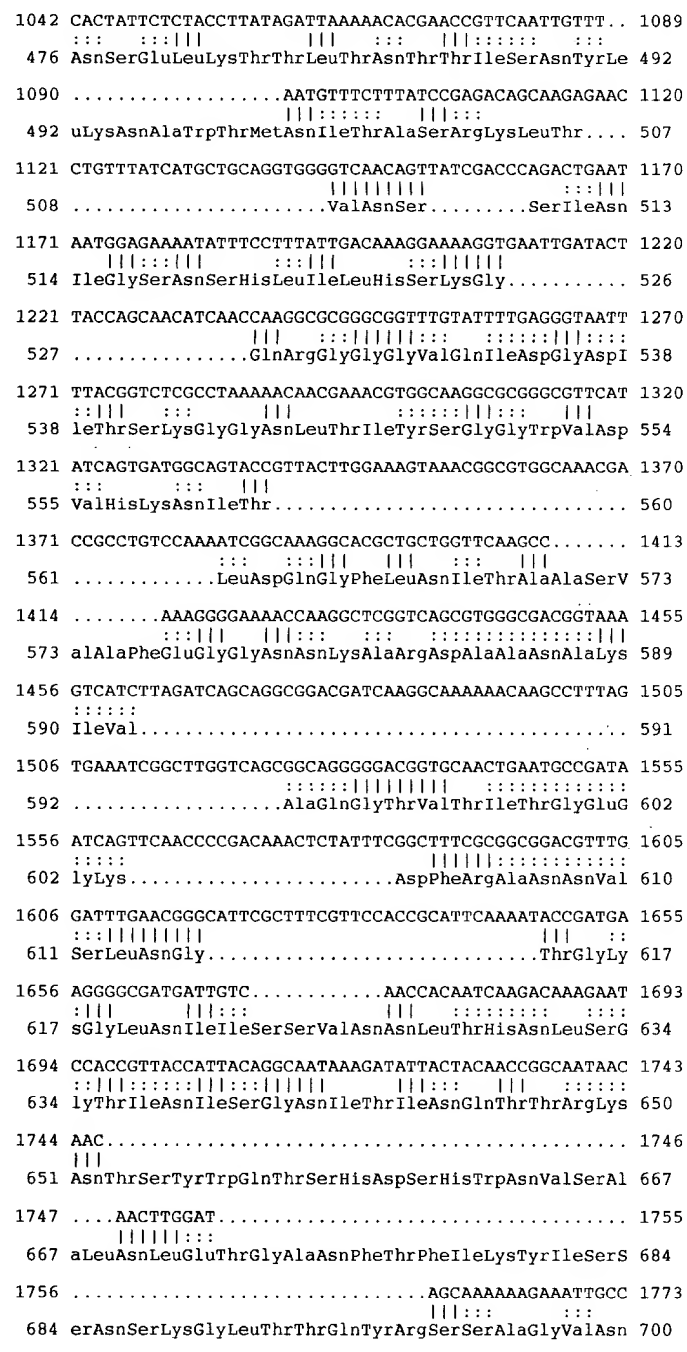
```
alignment_block:
 US-09-303-518D-653 x US-08-469-880-4   ..

 Align seg 1/1  to: US-08-469-880-4  from: 1  to: 1477

    178 AATAAAGGCAAGTTTGCAGTCGGGGCGAAAGATATTGAGGTTTACAACAA 227
        |||   |||   :::::::||||||   ||||||   :::|||
    199 AsnHisGlyLeuIleThrValGly...LysAspGlySerVal..... 211

    228 AAAAGGGGAGTTGGTCGGCAAATCGATGACGAAAGCCCCGATGATTGATT 277
              :::|||:::|||   ::::::   :::
    212 .......AsnLeuIleGlyGlyLysValLysAsnGlu..... 221

    278 TTTCTGTGGTATCGCGTAACGGC......GTGGCGGCATTGGCGGGCGAT 321
          :::|||:::|||   ||||||      ::::::   |||||||||:::
    222 ..GlyValIleSerValAsnGlyGlySerIleSerLeuLeuAlaGlyGln 237

    322 CAA.................................TATATTGTGAG 335
        :::                                    |||   :::::
    238 LysIleThrIleSerAspIleIleAsnProThrIleThrTyrSerIleAl 254

    336 CGTGGCACATAACGGCGGCTATAACAATGTTGATTTTGGTGCGGAGGGAA 385
        ::::   :::|||   :::   |||      |||:::   |||:::|||:
    254 aAlaProGluAsnGluAlaValAsnLeuGlyAspIlePheAlaLysGlyG 271

    386 GCAAT.......................CCCGATCAGCACCGCTTTTCT 411
        ::|||                          :::|||   ::::::|||
    271 lyAsnIleAsnValArgAlaAlaThrIleArgAsnGlnGlyLysLeuSer 287

    412 TACCAAATTGTGAAAAGAAATAATTATAAAGCAGGGACTAACGGCCATCC 461
           :::   |||:::::::::      |||:::|||
    288 AlaAspSerValSerLysAsp......LysSerGly..... 297

    462 TTATGGCGGCGATTATCATATGCCGCGTTTGCACAAATTTGTCACAGATG 511
                                          ::::::|||   :::|
    298 ..............................AsnIleValLeuSerA 303

    512 CAGAACCTGTTGAG.................ATGACCAGTTATATGGAT 543
        ||:::      |||                 :::::::::      :::
    303 laLysGluGlyGluAlaGluIleGlyGlyValIleSerAlaGlnAsnGln 319

    544 GGGTGGAAATACGCTGATTTAAATAAATACCCTGATCGTGTT......CG 587
              |||   :::   |||           |||:::|||      ::
    320 GlnAlaLysGlyGlyLysLeuMetIleThrGlyAspLysValThrLeuLy 336

    588 AATCGGAGCA.........................GGCAGACAATATT 610
        :   ||||||                        |||:::   |||
    336 sThrGlyAlaValIleAspLeuSerGlyLysGluGlyGlyGluThrTyrL 353

    611 GGCGGTCTGATGAA...GACGAACCCAATAAC.................. 639
              :::||||||      |||   :::|||
    353 euGlyGlyAspGluArgGlyGluGlyLysAsnGlyIleGlnLeuAlaLys 369

    640 ..............CGCGAAAGTTCATATCATATTGCAAGC........ 666
                        :::   |||:::   ::::::::::::
    370 LysThrSerLeuGluLysGlySerThrIleAsnValSerGlyLysGluLy 386

    667 ..........GCATATTCTTGG............CTCGTCGGTGGCAATA 694
                   |||      |||           |||:::   ||||||
    386 sGlyGlyPheAlaIleValTrpGlyAspIleAlaLeuIleAspGlyAsnI 403

    695 CCTTTGCACAAAATGGATCAGGTGGTGGCACAGTCAACTTAGGTAGCGAA 744
              ||||||   |||||||||
    403 leAsnAlaGln...GlySerGly.......................... 409

    745 AAAATTAAACATAGCCCATATGGTTTTTTACCAACAGGAGGCTCATTTGG 794
                                      :::      |||||||||
    410 ....................AspIleAlaLysThrGlyGlyPheValGl 419

    795 CGACAGTGGCTCACCAATGTTTATCTATGATGCCCAAAAGCAAAAGTGGT 844
           ||||||      :::||||||   |||
    419 uThrSerGlyHisAspLeuPheIleLysAsp.................. 429
```

```
 845 TAATTAATGGGGTATTGCAAACAGGCAACCCCTATATAGGAAAAAGCAAT 894
 430 .....AsnAlaIleValAspAla..................................... 435

 895 GGCTTCCAGCTAGTTCGTAAAGATTGGTTCTATGATGAAATCTTTGCTGG 944
 436 ..................LysGluTrpLeuLeuAsp............Ph 442

 945 AGATACCCATTCAGTATTCTACGAA...CCACATCAAAATGGGAAATACT 991
 442 eAspAsnValSerIleAsnAlaGluAspProLeuPheAsnAsnThrGlyI 459

 992 TTTTTAACGACAATAATAATGGCGCAGGAAAAATCGATGCCAAACATAAA 1041
 459 leAsnAspGluPheProThrGlyThrGlyGluAlaSerAspProLysLys 475

1042 CACTATTCTCTACCTTATAGATTAAAAACACGAACCGTTCAATTGTTT.. 1089
 476 AsnSerGluLeuLysThrThrLeuThrAsnThrThrIleSerAsnTyrLe 492

1090 ..................AATGTTTCTTTATCCGAGACAGCAAGAGAAC 1120
 492 uLysAsnAlaTrpThrMetAsnIleThrAlaSerArgLysLeuThr.... 507

1121 CTGTTTATCATGCTGCAGGTGGGGTCAACAGTTATCGACCCAGACTGAAT 1170
 508 ......................ValAsnSer.........SerIleAsn 513

1171 AATGGAGAAAATATTTCCTTTATTGACAAAGGAAAAGGTGAATTGATACT 1220
 514 IleGlySerAsnSerHisLeuIleLeuHisSerLysGly........... 526

1221 TACCAGCAACATCAACCAAGGCGCGGGCGGTTTGTATTTTGAGGGTAATT 1270
 527 ..............GlnArgGlyGlyGlyValGlnIleAspGlyAspI 538

1271 TTACGGTCTCGCCTAAAAACAACGAAACGTGGCAAGGCGCGGGCGTTCAT 1320
 538 leThrSerLysGlyGlyAsnLeuThrIleTyrSerGlyGlyTrpValAsp 554

1321 ATCAGTGATGGCAGTACCGTTACTTGGAAAGTAAACGGCGTGGCAAACGA 1370
 555 ValHisLysAsnIleThr.............................. 560

1371 CCGCCTGTCCAAAATCGGCAAAGGCACGCTGCTGGTTCAAGCC....... 1413
 561 ............LeuAspGlnGlyPheLeuAsnIleThrAlaAlaServ 573

1414 ........AAAGGGGAAAACCAAGGCTCGGTCAGCGTGGGCGACGGTAAA 1455
 573 alAlaPheGluGlyGlyAsnAsnLysAlaArgAspAlaAlaAsnAlaLys 589

1456 GTCATCTTAGATCAGCAGGCGGACGATCAAGGCAAAAAACAAGCCTTTAG 1505
 590 IleVal........................................... 591

1506 TGAAATCGGCTTGGTCAGCGGCAGGGGGACGGTGCAACTGAATGCCGATA 1555
 592 ..................AlaGlnGlyThrValThrIleThrGlyGluG 602

1556 ATCAGTTCAACCCCGACAAACTCTATTTCGGCTTTCGCGGCGGACGTTTG 1605
 602 lyLys.......................AspPheArgAlaAsnAsnVal 610

1606 GATTTGAACGGGCATTCGCTTTCGTTCCACCGCATTCAAAATACCGATGA 1655
 611 SerLeuAsnGly............................ThrGlyLy 617

1656 AGGGGCGATGATTGTC............AACCACAATCAAGACAAAGAAT 1693
 617 sGlyLeuAsnIleIleSerSerValAsnAsnLeuThrHisAsnLeuSerG 634

1694 CCACCGTTACCATTACAGGCAATAAAGATATTACTACAACCGGCAATAAC 1743
```

```
 634 lyThrIleAsnIleSerGlyAsnIleThrIleAsnGlnThrThrArgLys 650

1744 AAC.......................................... 1746
 651 AsnThrSerTyrTrpGlnThrSerHisAspSerHisTrpAsnValSerAl 667

1747 ....AACTTGGAT.................................. 1755
 667 aLeuAsnLeuGluThrGlyAlaAsnPheThrPheIleLysTyrIleSerS 684

1756 ............................AGCAAAAAAGAAATTGCC 1773
 684 erAsnSerLysGlyLeuThrThrGlnTyrArgSerSerAlaGlyValAsn 700

1774 TACAACGGTTGGTTTGGCGAGAAAGATGCAACCAAAACGAACGGGCGGCT 1823
 701 PheAsnGly.........................ValAsnGlyAsnMe 708

1824 CAATCTGAATTACCAACCGGAAGAAGCGGATCGCACTTTACTGCTTTCCG 1873
 708 tSerPheAsnLeuLys...GluGlyAlaLysValAsnPheLysLeuLysP 724

1874 GCGGAACAAATTTAAAC............................GGC 1893
 724 roAsnGluAsnMetAsnThrSerLysProLeuProIleArgPheLeuAla 740

1894 AATATCACGCAAACA...AACGGCAAACTGTTTTTCAGCGGCAGACCGAC 1940
 741 AsnIleThrAlaThrGlyGlyGlySerValPhePheAsp.......... 753

1941 ACCGCACGCCTAC...AATCATTTAGGAAGCGGGTGGTCAAAAATGGAAG 1987
 754 .......IleTyrAlaAsnHisSerGlyArgGly............AlaG 764

1988 GTATCCCACAAGGAGAAATCGTGTGGGACAACGATTGGATCGACCGCACA 2037
 764 luLeuLysMetSerGluIleAsnIleSerAsn................. 774

2038 TTTAAAGCGGAAAACTTC............CATATTCAGGGCGGACAAGC 2075
 775 ......GlyAlaAsnPheThrLeuAsnSerHisValArgGlyAspAspAl 789

2076 GGTGGTTTCCCGCAATGTTGCCAAAGTGGAAGGCGATTGGCATTTAAGCA 2125
 789 aPhe..................LysIleAsnLysAspLeuThrIleAsn. 799

2126 ATCACGCCCAAGCAGTTTTCGGTGTCGCACCGCATCAAAGCCACACAATC 2175

 799 ..................................................... 799

2176 TGTACACGTTCGGACTGGACGGGTCTGACAAGTTGTACCGAAAAAACCAT 2225
 800 AlaThrAsnSerAsnPhe............................... 805

2226 TACCGACGATAAAGTGATTGCTTCATTGAGCAAGACCGACATCAGAGGCA 2275
 806 .....................SerLeuArgGlnThrLysAspAspPheT 815

2276 ATGTCAGCCTTGCCGATCACGCTCATTTAAATCTCACAGGACTTGCCACA 2325
 815 yrAspGlyTyrAlaArgAsnAlaIleAsnSerThrTyrAsnIleSerIle 831

2326 CTCAACGGCAATCTTAGTGCAGGCGGAGACACGCACTATACGGTTACGCG 2375
 832 LeuGlyGlyAsnValThrLeuGlyGly...................... 840

2376 CAACGCCACCCAAAACGGCAACCTCAGCCTCGTGGGCAATGCCCAAGCAA 2425
 841 ..........GlnAsnSerSerSerSerIleThrGlyAsn......IleT 852

2426 CATTTAATCAAGCC.........ACATTAAACGGCAACACATCGGCTTCG 2466
```


852 hrIleGluLysAlaAlaAsnValThrLeuGluAlaAsnAsnAlaProAsn 868
2467 GACAATGCTTCATTTAATCTAAGCAACAACGCCGTACAAAACGGCAGTCT 2516
869 GlnGln.....AsnIleArgAspArgValIleLysLeuGlySerLe 882
2517 GACGCTTTCCGACAACGCTAAGGCAAACGTAAGCCATTCCGCACTCAACG 2566
882 uLeuValAsnGlySerLeuSerLeuThrGlyGluAsnAlaAspIleLysG 899
2567 GCAATGTCTCCCTAGCCGATAAGGCAGTATTCCAT.....TTTGAA 2607
899 lyAsnLeuThrIleSerGluSerAlaThrPheLysGlyLysThrArgAsp 915
2608 AACAGCCGCTTTACCGGAAAAATCAGCGGCGGCAAGGATACGGCATTACA 2657
916 ThrLeuAsnIleThrGlyAsnPheThrAsnAsnGlyThrAlaGluIleAs 932
2658 CTTAAAAGACAGCGAATGGACGCTGCCGTCGGGCACGGAATTAGGCAATT 2707
932 nIleThrGlnGlyValValLysLeuGlyAsnValThrAsnAspGlyAspL 949
2708 TAAACCTTGACAACGCCACCATTACACTCAATTCCGCCTATCGACACGAT 2757
949 euAsn.....IleThrThrHisAla..... 955
2758 GCGGCAGGCGCGCAAACCGGCAGTGCGGCAGATGCGCCGCGCCGCCGTTC 2807
956 .....LysArgAs 958
2808 GCGCCGTTCCCTATTATCCGTTACGCCGCCAACTTCGGCAGAATCCCGTT 2857
958 nGlnArgSerIle..... 962
2858 TCAACACGCTGACGGTAAACGGCAAATTGAACGGTCAGGGAACATTCCGC 2907
963 .....IleGlyGlyAspIleIleAsn..... 969
2908 TTTATGTCGGAACTCTTCGGCTACCGCAGCGGCAAATTGAAGCTGGCGGA 2957
970 .....LysLysGlySerLeuAsnIleThrAs 978
2958 AAGTTCC.....G 2965
978 pSerAsnAsnAspAlaGluIleGlnIleGlyGlyAsnIleSerGlnLysG 995
2966 AAGGCACTTACACCTTGGCT.....GTCAACAATACCGGCAACGAA 3006
995 luGlyAsnLeuThrIleSerSerAspLysIleAsnIleThr..... 1008
3007 CCCGTAAGTCTCGAGCAATTGACGGTA.....GTGGAAGGAAAAGA 3047
1009 .....LysGlnIleThrIleLysLysGlyIleAspGlyGluAs 1021
3048 C...AACACACCGCTGTCCGAAAATCTTAATTTCACCCTGCAAAACGAAC 3094
1021 pSerSerSerAspAlaThrSerAsnAlaAsnLeuThrIleLys..... 1035
3095 ACGTCGATGCCGGCGCATGGCGTTATCAGCTTATCCGCAAAGACGGCGAG 3144
1035 1035
3145 TTCCGCCTGCATAATCCGGTCAAAGAACAAGAGCTTTCCGACAAACTCGG 3194
1036 .....ThrLysGluLeuLysLeuThrGluAspLeuSe 1046
3195 CAAGGCGGGAGAAACAGAGGCCGCCTTGACGGCAAAACAGGCACAACTTG 3244
1046 rIleSerGlyPheAsnLysAlaGluIleThrAlaLysAspGlyArgAspL 1063
3245 CCGCCAAACAACAGGCGGAAAAAGACAACGCGCAAAGC..... 3282
1063 euThrIleGlyAsnSerAsnAspGlyAsnSerGlyAlaGluAlaLysThr 1079
3283 .....CTTGACGCGCTGATTGCGGCC...GGGCGCAA 3311
1080 ValThrPheAsnAsnValLysAspSerLysIleSerAlaAspGlyHisAs 1096
3312 TGCCACCGAAAAGGCAGAAAGTGTTGCCGAACCGGCCCGGCAGGCAGGCG 3361
1096 nValThr.....LeuAsnSerLysValLysThrSerSerSerAsnGlyG 1111
3362 GGGAAAATGCCGGCATTATGCAGGCGGAGGAAGAGAAAAAACGGGTGCAG 3411
1111 lyArgGluSer.....Asn 1115
3412 GCGGATAAAGACACCGCCTTG.....GCGAAACAGCGCGAAGCGGA 3452
1116 SerAspAsnAspThrGlyLeuThrIleThrAlaLysAsnValGluVal.. 1131
3453 AACCCGGCCGGCTACCACCGCCTTCCCCCGCGCCCGCCGCGCCCGCCGGG 3502
1131 1131
3503 ATTTGCCGCAACCGCAGCCCCAACCGCAACCCCAACCGCAGCGCGACCTG 3552
1132 .....AsnLysAspIle 1135
3553 ATCAGC...CGTTATGCCAATAGCGGTTTGAGTGAATTTTCCGCCACGCT 3599
1136 ThrSerLeuLysThrValAsnIleThrAlaSerGluLysValThrThr.T 1152
3600 CAACAGCGTTTTCGCCGTACAGGACGAATTGGACCGCGTGTTTGCCGAAG 3649
1152 hrAlaGlySerThrIleAsnAlaThrAsnGlyLysAlaSerIleThrThr 1168
3650 ACCGCCGCAACGCCGTTTGGACAAGCGGCATCCGGGACACCAAACACTAC 3699
1169 LysThrGlyAspIleSerGlyThrIleSerGlyAsnThrValSerValSe 1185
3700 CGTTCGC...AAGATTTCCGCGCCTACCGCCAACAAACCGACCTGCGCCA 3746
1185 rAlaThrValAspLeuThrThrLysSerGlySerLysIleGluAlaLysS 1202
3747 AATCGGTATGCAGAAAAACCTCGGCAGCGGGCGCGTCGGCATCCTGTTTT 3796
1202 erGlyGluAlaAsnValThrSerAlaThrGlyThrIleGlyGly..... 1216
3797 CGCACAACCGGACCGGAAACACCTTCGACGACGGCATCGGCA..... 3838
1217 ...ThrIleSerGlyAsnThrValAsnValThrAlaAsnAlaGlyAspLe 1232
3839 .....ACTCGGCACGGCTTGCCCACGG 3860
1232 uThrValGlyAsnGlyAlaGluIleAsnAlaThrGluGlyAlaAlaThrL 1249
3861 TGCCGTTTTCGGGCAAT......ACGGCATCGGCAGGTTCGACATCGGCA 3904
1249 euThrAlaThrGlyAsnThrLeuThrThrGluAlaGlySerSerIleThr 1265
3905 TCAGCGCGGGCGCGGGTTTTAGTAGCGGCAGCCTTTCAGACGGCATCAGA 3954
1266 SerThrLysGlyGlnValAspLeuLeuAla......GlnAsnGlySerIl 1280
3955 GGCAAAATCCGCCGCCGCGTGCTGCATTACGGCATTCAGGCAAGATACCG 4004
1280 eAlaGlySerIleAsnAlaAlaAsnValThrLeuAsnThrThrGlyThrL 1297
4005 CGCAGGTTTCGGCGGATTCGGCATCGAACCGCACATCGGCGCAACGCGCT 4054
1297 euThrThrValAlaGlySerAspIleLysAlaThrSerGlyThrLeuVal 1313
4055 ATTTCGTCCAAA 4066
1314 IleAsnAlaLys 1317

```
seq_name: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:US-08-728-470-4

seq_documentation_block:
; Sequence 4, Application US/08728470
; Patent No. 5928651
;  GENERAL INFORMATION:
;    APPLICANT:  Barenkamp, Stephen J
;    TITLE OF INVENTION:  High Molecular Weight Surface Proteins
;    TITLE OF INVENTION:  of No. 5928651-Typeable Haemophilus
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Shoemaker and Mattare, Ltd.
;      STREET:  2001 Jefferson Davis Hwy., 1203 Crystal Plaza
;      STREET:  Bldg. 1
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22202-0286
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/728,470
;      FILING DATE:
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/302,832
;      FILING DATE:  16-MAR-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US PCT/US93/02166
;      FILING DATE:  16-MAR-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9205704.1
;      FILING DATE:  16-MAR-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Berkstresser, Jerry W
;      REGISTRATION NUMBER:  22,651
;      REFERENCE/DOCKET NUMBER:  1038-633
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 415-0810
;      TELEFAX:  (703) 415-0813
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1477 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-728-470-4


alignment_scores:
            Quality:  274.50              Length:   1488
              Ratio:  0.419                 Gaps:     73
 Percent Similarity:  44.019   Percent Identity: 19.220

alignment_block:
 US-09-303-518D-653 x US-08-728-470-4    ..

 Align seg 1/1  to: US-08-728-470-4  from: 1  to: 1477

     178 AATAAAGGCAAGTTTGCAGTCGGGGCGAAAGATATTGAGGTTTACAACAA 227
         |||   |||   :::::::||||||   ||||||   :::|||
     199 AsnHisGlyLeuIleThrValGly...LysAspGlySerVal..... 211

     228 AAAAGGGGAGTTGGTCGGCAAATCGATGACGAAAGCCCCGATGATTGATT 277
               :::|||:::|||   ::::::   :::
     212 .......AsnLeuIleGlyGlyLysValLysAsnGlu..... 221

     278 TTTCTGTGGTATCGCGTAACGGC......GTGGCGGCATTGGCGGGCGAT 321
           :::|||:::|||   ||||||      ::::::   |||||||||:::
     222 ..GlyValIleSerValAsnGlyGlySerIleSerLeuLeuAlaGlyGln 237

     322 CAA.................................TATATTGTGAG 335
         :::                                 |||   :::::
     238 LysIleThrIleSerAspIleIleAsnProThrIleThrTyrSerIleAl 254

     336 CGTGGCACATAACGGCGGCTATAACAATGTTGATTTTGGTGCGGAGGGAA 385
         ::::   :::|||   :::   |||      |||:::   |||:::|||:
     254 aAlaProGluAsnGluAlaValAsnLeuGlyAspIlePheAlaLysGlyG 271

     386 GCAAT.......................CCCGATCAGCACCGCTTTTCT 411
         ::|||                       :::|||   ::::::|||
     271 lyAsnIleAsnValArgAlaAlaThrIleArgAsnGlnGlyLysLeuSer 287

     412 TACCAAATTGTGAAAAGAAATAATTATAAAGCAGGGACTAACGGCCATCC 461
            :::   |||:::::::::      |||:::|||
     288 AlaAspSerValSerLysAsp......LysSerGly..... 297

     462 TTATGGCGGCGATTATCATATGCCGCGTTTGCACAAATTTGTCACAGATG 511
                                       ::::::|||   :::|
     298 ................................AsnIleValLeuSerA 303

     512 CAGAACCTGTTGAG.................ATGACCAGTTATATGGAT 543
         ||:::      |||                 :::::::::      :::
     303 laLysGluGlyGluAlaGluIleGlyGlyValIleSerAlaGlnAsnGln 319

     544 GGGTGGAAATACGCTGATTTAAATAAATACCCTGATCGTGTT......CG 587
               |||   :::   |||            |||:::|||      ::
     320 GlnAlaLysGlyGlyLysLeuMetIleThrGlyAspLysValThrLeuLy 336

     588 AATCGGAGCA..........................GGCAGACAATATT 610
         :   ||||||                         |||:::   |||
     336 sThrGlyAlaValIleAspLeuSerGlyLysGluGlyGlyGluThrTyrL 353

     611 GGCGGTCTGATGAA...GACGAACCCAATAAC..................... 639
               :::||||||      |||   :::|||
     353 euGlyGlyAspGluArgGlyGluGlyLysAsnGlyIleGlnLeuAlaLys 369

     640 ..............CGCGAAAGTTCATATCATATTGCAAGC........ 666
                      :::   |||:::   ::::::::::::
     370 LysThrSerLeuGluLysGlySerThrIleAsnValSerGlyLysGluLy 386

     667 ..........GCATATTCTTGG............CTCGTCGGTGGCAATA 694
                    |||      |||           |||:::   ||||||
     386 sGlyGlyPheAlaIleValTrpGlyAspIleAlaLeuIleAspGlyAsnI 403

     695 CCTTTGCACAAAATGGATCAGGTGGTGGCACAGTCAACTTAGGTAGCGAA 744
               ||||||   |||||||||
     403 leAsnAlaGln...GlySerGly.......................... 409

     745 AAAATTAAACATAGCCCATATGGTTTTTTACCAACAGGAGGCTCATTTGG 794
                                 :::      |||||||||
     410 ....................AspIleAlaLysThrGlyGlyPheValGl 419

     795 CGACAGTGGCTCACCAATGTTTATCTATGATGCCCAAAAGCAAAAGTGGT 844
            ||||||      :::||||||   |||
     419 uThrSerGlyHisAspLeuPheIleLysAsp.................. 429

     845 TAATTAATGGGGTATTGCAAACAGGCAACCCCTATATAGGAAAAAGCAAT 894
               |||:::::::::::::::
     430 .....AsnAlaIleValAspAla.......................... 435

     895 GGCTTCCAGCTAGTTCGTAAAGATTGGTTCTATGATGAAATCTTTGCTGG 944
                         |||:::|||:::   |||
     436 ..................LysGluTrpLeuLeuAsp............Ph 442

     945 AGATACCCATTCAGTATTCTACGAA...CCACATCAAAATGGGAAATACT 991
          |||:::   |||:::      |||   |||      |||:::      :
     442 eAspAsnValSerIleAsnAlaGluAspProLeuPheAsnAsnThrGlyI 459

     992 TTTTTAACGACAATAATAATGGCGCAGGAAAAATCGATGCCAAACATAAA 1041
         ::   ::::::      :::|||:::|||:::   :::          |||
     459 leAsnAspGluPheProThrGlyThrGlyGluAlaSerAspProLysLys 475
```


1042 CACTATTCTCTACCTTATAGATTAAAAACACGAACCGTTCAATTGTTT.. 1089
476 AsnSerGluLeuLysThrThrLeuThrAsnThrThrIleSerAsnTyrLe 492
1090 ..................AATGTTTCTTTATCCGAGACAGCAAGAGAAC 1120
492 uLysAsnAlaTrpThrMetAsnIleThrAlaSerArgLysLeuThr.... 507
1121 CTGTTTATCATGCTGCAGGTGGGGTCAACAGTTATCGACCCAGACTGAAT 1170
508 .....ValAsnSer.....SerIleAsn 513
1171 AATGGAGAAAATATTTCCTTTATTGACAAAGGAAAAGGTGAATTGATACT 1220
514 IleGlySerAsnSerHisLeuIleLeuHisSerLysGly..... 526
1221 TACCAGCAACATCAACCAAGGCGCGGGCGGTTTGTATTTTGAGGGTAATT 1270
527 .....GlnArgGlyGlyGlyValGlnIleAspGlyAspI 538
1271 TTACGGTCTCGCCTAAAAACAACGAAACGTGGCAAGGCGCGGGCGTTCAT 1320
538 leThrSerLysGlyGlyAsnLeuThrIleTyrSerGlyGlyTrpValAsp 554
1321 ATCAGTGATGGCAGTACCGTTACTTGGAAAGTAAACGGCGTGGCAAACGA 1370
555 ValHisLysAsnIleThr..... 560
1371 CCGCCTGTCCAAAATCGGCAAAGGCACGCTGCTGGTTCAAGCC..... 1413
561 .....LeuAspGlnGlyPheLeuAsnIleThrAlaAlaSerV 573
1414 .....AAAGGGGAAAACCAAGGCTCGGTCAGCGTGGGCGACGGTAAA 1455
573 alAlaPheGluGlyGlyAsnAsnLysAlaArgAspAlaAlaAsnAlaLys 589
1456 GTCATCTTAGATCAGCAGGCGGACGATCAAGGCAAAAAACAAGCCTTTAG 1505
590 IleVal..... 591
1506 TGAAATCGGCTTGGTCAGCGGCAGGGGGACGGTGCAACTGAATGCCGATA 1555
592 .....AlaGlnGlyThrValThrIleThrGlyGluG 602
1556 ATCAGTTCAACCCCGACAAACTCTATTTCGGCTTTCGCGGCGGACGTTTG 1605
602 lyLys.....AspPheArgAlaAsnAsnVal 610
1606 GATTTGAACGGGCATTCGCTTTCGTTCCACCGCATTCAAAATACCGATGA 1655
611 SerLeuAsnGly.....ThrGlyLy 617
1656 AGGGGCGATGATTGTC.....AACCACAATCAAGACAAAGAAT 1693
617 sGlyLeuAsnIleIleSerSerValAsnAsnLeuThrHisAsnLeuSerG 634
1694 CCACCGTTACCATTACAGGCAATAAAGATATTACTACAACCGGCAATAAC 1743
634 lyThrIleAsnIleSerGlyAsnIleThrIleAsnGlnThrThrArgLys 650
1744 AAC..... 1746
651 AsnThrSerTyrTrpGlnThrSerHisAspSerHisTrpAsnValSerAl 667
1747 ....AACTTGGAT..... 1755
667 aLeuAsnLeuGluThrGlyAlaAsnPheThrPheIleLysTyrIleSerS 684
1756 .....AGCAAAAAAGAAATTGCC 1773
684 erAsnSerLysGlyLeuThrThrGlnTyrArgSerSerAlaGlyValAsn 700
1774 TACAACGGTTGGTTTGGCGAGAAAGATGCAACCAAAACGAACGGGCGGCT 1823
701 PheAsnGly.....ValAsnGlyAsnMe 708
1824 CAATCTGAATTACCAACCGGAAGAAGCGGATCGCACTTTACTGCTTTCCG 1873
708 tSerPheAsnLeuLys...GluGlyAlaLysValAsnPheLysLeuLysP 724
1874 GCGGAACAAATTTAAAC.....GGC 1893
724 roAsnGluAsnMetAsnThrSerLysProLeuProIleArgPheLeuAla 740
1894 AATATCACGCAAACA...AACGGCAAACTGTTTTTCAGCGGCAGACCGAC 1940
741 AsnIleThrAlaThrGlyGlyGlySerValPhePheAsp..... 753
1941 ACCGCACGCCTAC...AATCATTTAGGAAGCGGGTGGTCAAAAATGGAAG 1987
754 .....IleTyrAlaAsnHisSerGlyArgGly.....AlaG 764
1988 GTATCCCACAAGGAGAAATCGTGTGGGACAACGATTGGATCGACCGCACA 2037
764 luLeuLysMetSerGluIleAsnIleSerAsn..... 774
2038 TTTAAAGCGGAAAACTTC.....CATATTCAGGGCGGACAAGC 2075
775 .....GlyAlaAsnPheThrLeuAsnSerHisValArgGlyAspAspAl 789
2076 GGTGGTTTCCCGCAATGTTGCCAAAGTGGAAGGCGATTGGCATTTAAGCA 2125
789 aPhe.....LysIleAsnLysAspLeuThrIleAsn. 799
2126 ATCACGCCCAAGCAGTTTTCGGTGTCGCACCGCATCAAAGCCACACAATC 2175
799 799
2176 TGTACACGTTCGGACTGGACGGGTCTGACAAGTTGTACCGAAAAAACCAT 2225
800 AlaThrAsnSerAsnPhe..... 805
2226 TACCGACGATAAAGTGATTGCTTCATTGAGCAAGACCGACATCAGAGGCA 2275
806 .....SerLeuArgGlnThrLysAspAspPheT 815
2276 ATGTCAGCCTTGCCGATCACGCTCATTTAAATCTCACAGGACTTGCCACA 2325
815 yrAspGlyTyrAlaArgAsnAlaIleAsnSerThrTyrAsnIleSerIle 831
2326 CTCAACGGCAATCTTAGTGCAGGCGGAGACACGCACTATACGGTTACGCG 2375
832 LeuGlyGlyAsnValThrLeuGlyGly..... 840
2376 CAACGCCACCCAAAACGGCAACCTCAGCCTCGTGGGCAATGCCCAAGCAA 2425
841 .....GlnAsnSerSerSerSerIleThrGlyAsn.....IleT 852
2426 CATTTAATCAAGCC.....ACATTAAACGGCAACACATCGGCTTCG 2466
852 hrIleGluLysAlaAlaAsnValThrLeuGluAlaAsnAsnAlaProAsn 868
2467 GACAATGCTTCATTTAATCTAAGCAACAACGCCGTACAAAACGGCAGTCT 2516
869 GlnGln.....AsnIleArgAspArgValIleLysLeuGlySerLe 882
2517 GACGCTTTCCGACAACGCTAAGGCAAACGTAAGCCATTCCGCACTCAACG 2566
882 uLeuValAsnGlySerLeuSerLeuThrGlyGluAsnAlaAspIleLysG 899
2567 GCAATGTCTCCCTAGCCGATAAGGCAGTATTCCAT.....TTTGAA 2607
899 lyAsnLeuThrIleSerGluSerAlaThrPheLysGlyLysThrArgAsp 915
2608 AACAGCCGCTTTACCGGAAAAATCAGCGGCGGCAAGGATACGGCATTACA 2657

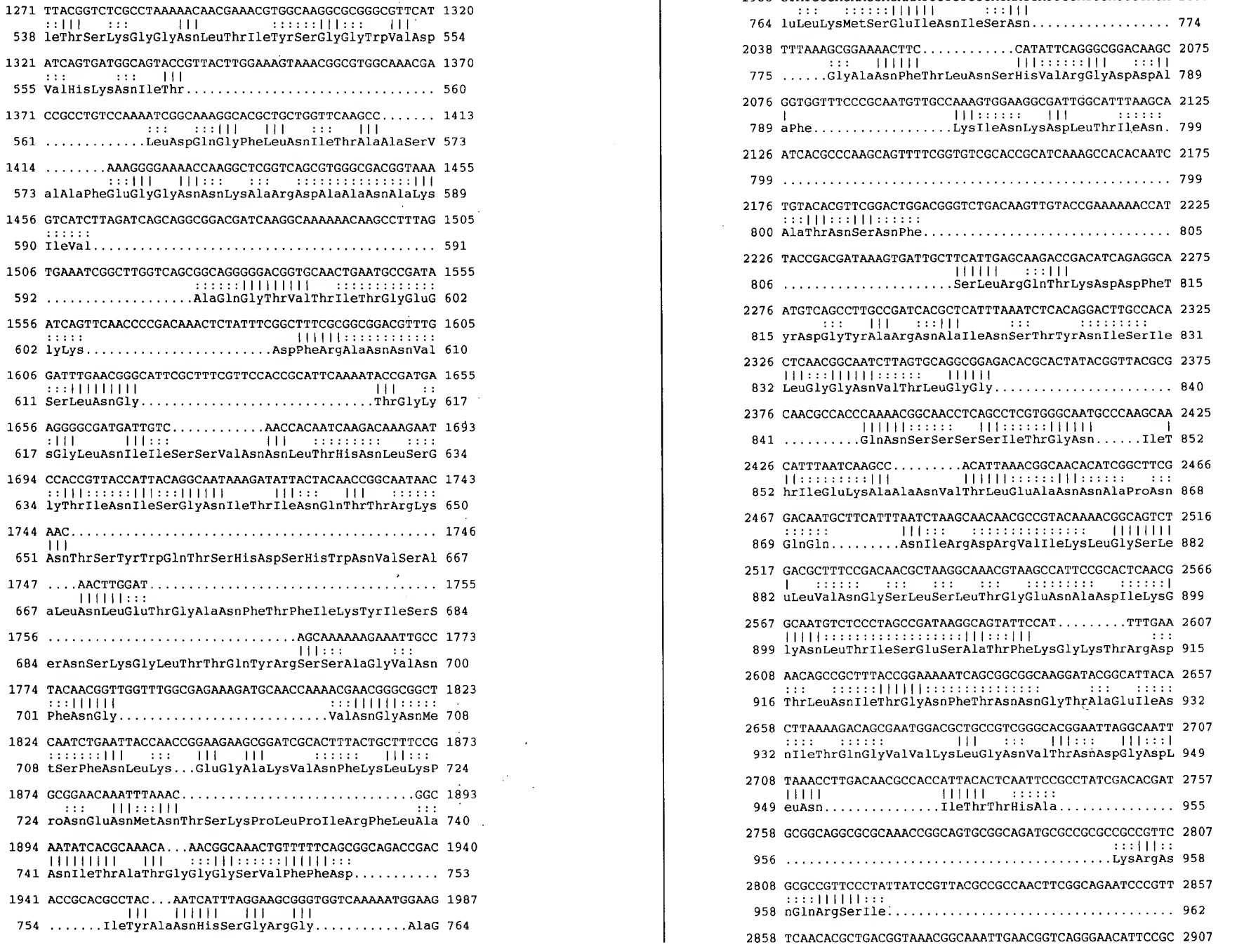
```
 916 ThrLeuAsnIleThrGlyAsnPheThrAsnAsnGlyThrAlaGluIleAs 932

2658 CTTAAAAGACAGCGAATGGACGCTGCCGTCGGGCACGGAATTAGGCAATT 2707
     ::::   ::::::        |||   :::   |||:::   |||:::|
 932 nIleThrGlnGlyValValLysLeuGlyAsnValThrAsnAspGlyAspL 949

2708 TAAACCTTGACAACGCCACCATTACACTCAATTCCGCCTATCGACACGAT 2757
     |||||               ||||||   ::::::
 949 euAsn..............IleThrThrHisAla.............. 955

2758 GCGGCAGGCGCGCAAACCGGCAGTGCGGCAGATGCGCCGCGCCGCCGTTC 2807
                                                 :::|||::
 956 .....................................LysArgAs 958

2808 GCGCCGTTCCCTATTATCCGTTACGCCGCCAACTTCGGCAGAATCCCGTT 2857
     ::::|||||||:::
 958 nGlnArgSerIle.................................... 962

2858 TCAACACGCTGACGGTAAACGGCAAATTGAACGGTCAGGGAACATTCCGC 2907
                  :::::::|||   :::   :::
 963 ..............IleGlyGlyAspIleIleAsn.............. 969

2908 TTTATGTCGGAACTCTTCGGCTACCGCAGCGGCAAATTGAAGCTGGCGGA 2957
                            :::::::|||:::|||::::::::::::
 970 .......................LysLysGlySerLeuAsnIleThrAs 978

2958 AAGTTCC.........................................G 2965
     :|||:::                                        |
 978 pSerAsnAsnAspAlaGluIleGlnIleGlyGlyAsnIleSerGlnLysG 995

2966 AAGGCACTTACACCTTGGCT.........GTCAACAATACCGGCAACGAA 3006
     ||||||::   |||::::::         :::|||   |||
 995 luGlyAsnLeuThrIleSerSerAspLysIleAsnIleThr......... 1008

3007 CCCGTAAGTCTCGAGCAATTGACGGTA.........GTGGAAGGAAAAGA 3047
                :::|||:::|||:::         ::::::|||:::||
1009 ............LysGlnIleThrIleLysLysGlyIleAspGlyGluAs 1021

3048 C...AACACACCGCTGTCCGAAAATCTTAATTTCACCCTGCAAAACGAAC 3094
     |   ::::::      ::::::|||   |||:::|||::::::
1021 pSerSerSerAspAlaThrSerAsnAlaAsnLeuThrIleLys....... 1035

3095 ACGTCGATGCCGGCGCATGGCGTTATCAGCTTATCCGCAAAGACGGCGAG 3144

1035 ................................................. 1035

3145 TTCCGCCTGCATAATCCGGTCAAAGAACAAGAGCTTTCCGACAAACTCGG 3194
                    :::|||||||   :::|||::::::   |||::
1036 ..................ThrLysGluLeuLysLeuThrGluAspLeuSe 1046

3195 CAAGGCGGGAGAAACAGAGGCCGCCTTGACGGCAAAACAGGCACAACTTG 3244
     :   :::|||   ::::::|||   :::||||||||:::::::::
1046 rIleSerGlyPheAsnLysAlaGluIleThrAlaLysAspGlyArgAspL 1063

3245 CCGCCAAACAACAGGCGGAAAAAGACAACGCGCAAAGC............ 3282
       :::      :::::::::      |||:::   :::
1063 euThrIleGlyAsnSerAsnAspGlyAsnSerGlyAlaGluAlaLysThr 1079

3283 ................CTTGACGCGCTGATTGCGGCC...GGGCGCAA 3311
                     |||:::   |||:::|||   |||:::|||
1080 ValThrPheAsnAsnValLysAspSerLysIleSerAlaAspGlyHisAs 1096

3312 TGCCACCGAAAAGGCAGAAAGTGTTGCCGAACCGGCCCGGCAGGCAGGCG 3361
     |:::|||         :::|||   ::::::   :::   :::   ||||
1096 nValThr......LeuAsnSerLysValLysThrSerSerSerAsnGlyG 1111

3362 GGGAAAATGCCGGCATTATGCAGGCGGAGGAAGAGAAAAAACGGGTGCAG 3411
     ||:::::::::                                    :::
1111 lyArgGluSer.................................Asn 1115

3412 GCGGATAAAGACACCGCCTTG.........GCGAAACAGCGCGAAGCGGA 3452
     :::|||:::||||||:::|||         ||||||:::   |||:::
1116 SerAspAsnAspThrGlyLeuThrIleThrAlaLysAsnValGluVal.. 1131
```

```
3453 AACCCGGCCGGCTACCACCGCCTTCCCCCGCGCCCGCCGCGCCCGCCGGG 3502

1131 ................................................. 1131

3503 ATTTGCCGCAACCGCAGCCCCAACCGCAACCCCAACCGCAGCGCGACCTG 3552
                                                 :::::::|||:::
1132 .....................................AsnLysAspIle 1135

3553 ATCAGC...CGTTATGCCAATAGCGGTTTGAGTGAATTTTCCGCCACGCT 3599
        |||   :::   :::|||         ||||||      :::||| :
1136 ThrSerLeuLysThrValAsnIleThrAlaSerGluLysValThrThr.T 1152

3600 CAACAGCGTTTTCGCCGTACAGGACGAATTGGACCGCGTGTTTGCCGAAG 3649
     ::::::::  :::       ||||||      |||   :::
1152 hrAlaGlySerThrIleAsnAlaThrAsnGlyLysAlaSerIleThrThr 1168

3650 ACCGCCGCAACGCCGTTTGGACAAGCGGCATCCGGGACACCAAACACTAC 3699
        ::::::         |||      ::::::::::|||   ::::::::
1169 LysThrGlyAspIleSerGlyThrIleSerGlyAsnThrValSerValSe 1185

3700 CGTTCGC...AAGATTTCCGCGCCTACCGCCAACAAACCGACCTGCGCCA 3746
     ::::         :::::::::   ::::::::::|||      :::   :
1185 rAlaThrValAspLeuThrThrLysSerGlySerLysIleGluAlaLysS 1202

3747 AATCGGTATGCAGAAAAACCTCGGCAGCGGGCGCGTCGGCATCCTGTTTT 3796
     :::::   ::::::   |||||||||:::|||:::   ::::::
1202 erGlyGluAlaAsnValThrSerAlaThrGlyThrIleGlyGly...... 1216

3797 CGCACAACCGGACCGGAAACACCTTCGACGACGGCATCGGCA........ 3838
        |||   :::   :::|||   ::::::||||||:::|||
1217 ...ThrIleSerGlyAsnThrValAsnValThrAlaAsnAlaGlyAspLe 1232

3839 ..........................ACTCGGCACGGCTTGCCCACGG 3860
                                    :::   :::|||      |||:
1232 uThrValGlyAsnGlyAlaGluIleAsnAlaThrGluGlyAlaAlaThrL 1249

3861 TGCCGTTTTCGGGCAAT......ACGGCATCGGCAGGTTCGACATCGGCA 3904
     ::      :::||||||      |||::::::||||||||:::   :::
1249 euThrAlaThrGlyAsnThrLeuThrThrGluAlaGlySerSerIleThr 1265

3905 TCAGCGCGGGCGCGGGTTTTAGTAGCGGCAGCCTTTCAGACGGCATCAGA 3954
     |||:::::::::::::|||   :::   |||      |||::::::|||
1266 SerThrLysGlyGlnValAspLeuLeuAla......GlnAsnGlySerIl 1280

3955 GGCAAAATCCGCCGCCGCGTGCTGCATTACGGCATTCAGGCAAGATACCG 4004
      |||   |||      |||:::   :::|||                |||
1280 eAlaGlySerIleAsnAlaAlaAsnValThrLeuAsnThrThrGlyThrL 1297

4005 CGCAGGTTTCGGCGGATTCGGCATCGAACCGCACATCGGCGCAACGCGCT 4054
          :::   |||   |||      :::   |||||||:::      :::
1297 euThrThrValAlaGlySerAspIleLysAlaThrSerGlyThrLeuVal 1313

4055 ATTTCGTCCAAA 4066
     |||:::::::|||
1314 IleAsnAlaLys 1317
```

seq_name: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:US-08-617-697-4

seq_documentation_block:
; Sequence 4, Application US/08617697
; Patent No. 5977336
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen J
; TITLE OF INVENTION: High Molecular Weight Surface Proteins
; TITLE OF INVENTION: of No. 5977336-Typeable Haemophilus
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Mattare, Ltd.
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; STREET: Bldg. 1
; CITY: Arlington
; STATE: Virginia

```
;       COUNTRY:  U.S.A.
;       ZIP:  22202-0286
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/617,697
;       FILING DATE:  01-APR-1996
;       CLASSIFICATION:  424
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/302,832
;       FILING DATE:  05-OCT-1994
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US PCT/US93/02166
;       FILING DATE:  16-MAR-1993
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Berkstresser, Jerry W
;       REGISTRATION NUMBER:  22,651
;       REFERENCE/DOCKET NUMBER:  1038-557
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (703) 415-0810
;       TELEFAX:  (703) 415-0813
;   INFORMATION FOR SEQ ID NO:  4:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1477 amino acids
;       TYPE:  amino acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
US-08-617-697-4


alignment_scores:
            Quality:  274.50            Length:   1488
              Ratio:  0.419               Gaps:     73
 Percent Similarity:  44.019   Percent Identity: 19.220

alignment_block:
 US-09-303-518D-653 x US-08-617-697-4   ..

 Align seq 1/1  to: US-08-617-697-4  from: 1  to: 1477

     178 AATAAAGGCAAGTTTGCAGTCGGGGCGAAAGATATTGAGGTTTACAACAA 227
         |||   |||   ::::::||||||   ||||||   :::|||
     199 AsnHisGlyLeuIleThrValGly...LysAspGlySerVal..... 211

     228 AAAAGGGGAGTTGGTCGGCAAATCGATGACGAAAGCCCCGATGATTGATT 277
               :::|||:::|||   ::::::   :::
     212 .......AsnLeuIleGlyGlyLysValLysAsnGlu..... 221

     278 TTTCTGTGGTATCGCGTAACGGC......GTGGCGGCATTGGCGGGCGAT 321
           :::|||:::|||   ||||||      ::::::   |||||||||:::
     222 ..GlyValIleSerValAsnGlyGlySerIleSerLeuLeuAlaGlyGln 237

     322 CAA.....................................TATATTGTGAG 335
         :::                                  |||   :::::
     238 LysIleThrIleSerAspIleIleAsnProThrIleThrTyrSerIleAl 254

     336 CGTGGCACATAACGGCGGCTATAACAATGTTGATTTTGGTGCGGAGGGAA 385
         ::::   :::|||   :::   |||      |||:::   |||:::|||:
     254 aAlaProGluAsnGluAlaValAsnLeuGlyAspIlePheAlaLysGlyG 271

     386 GCAAT.......................CCCGATCAGCACCGCTTTTCT 411
         ::|||                      :::|||   :::::::|||
     271 lyAsnIleAsnValArgAlaAlaThrIleArgAsnGlnGlyLysLeuSer 287

     412 TACCAAATTGTGAAAAGAAATAATTATAAAGCAGGGACTAACGGCCATCC 461
           :::   |||:::::::::      |||:::|||
     288 AlaAspSerValSerLysAsp......LysSerGly..... 297

     462 TTATGGCGGCGATTATCATATGCCGCGTTTGCACAAATTTGTCACAGATG 511
                                          ::::::|||   :::|

     298 ................................AsnIleValLeuSerA 303

     512 CAGAACCTGTTGAG.................ATGACCAGTTATATGGAT 543
         ||:::     |||                 :::::::::      :::
     303 laLysGluGlyGluAlaGluIleGlyGlyValIleSerAlaGlnAsnGln 319

     544 GGGTGGAAATACGCTGATTTAAATAAATACCCTGATCGTGTT......CG 587
               |||   :::   |||           |||:::|||      ::
     320 GlnAlaLysGlyGlyLysLeuMetIleThrGlyAspLysValThrLeuLy 336

     588 AATCGGAGCA.....................GGCAGACAATATT 610
         :   ||||||                     |||:::   |||
     336 sThrGlyAlaValIleAspLeuSerGlyLysGluGlyGlyGluThrTyrL 353

     611 GGCGGTCTGATGAA...GACGAACCCAATAAC..... 639
           :::||||||      |||   :::|||
     353 euGlyGlyAspGluArgGlyGluGlyLysAsnGlyIleGlnLeuAlaLys 369

     640 ..............CGCGAAAGTTCATATCATATTGCAAGC..... 666
                        :::   |||:::   ::::::::::::
     370 LysThrSerLeuGluLysGlySerThrIleAsnValSerGlyLysGluLy 386

     667 ..........GCATATTCTTGG............CTCGTCGGTGGCAATA 694
                 |||      |||            |||:::   ||||||
     386 sGlyGlyPheAlaIleValTrpGlyAspIleAlaLeuIleAspGlyAsnI 403

     695 CCTTTGCACAAAATGGATCAGGTGGTGGCACAGTCAACTTAGGTAGCGAA 744
           ||||||   |||||||||
     403 leAsnAlaGln...GlySerGly..... 409

     745 AAAATTAAACATAGCCCATATGGTTTTTTACCAACAGGAGGCTCATTTGG 794
                                     :::      |||||||||
     410 .....................AspIleAlaLysThrGlyGlyPheValGl 419

     795 CGACAGTGGCTCACCAATGTTTATCTATGATGCCCAAAAGCAAAAGTGGT 844
           ||||||      :::||||||   |||
     419 uThrSerGlyHisAspLeuPheIleLysAsp..... 429

     845 TAATTAATGGGGTATTGCAAACAGGCAACCCCTATATAGGAAAAAGCAAT 894
           |||:::::::::::::::
     430 .....AsnAlaIleValAspAla..... 435

     895 GGCTTCCAGCTAGTTCGTAAAGATTGGTTCTATGATGAAATCTTTGCTGG 944
                              |||:::|||:::   |||
     436 ..................LysGluTrpLeuLeuAsp............Ph 442

     945 AGATACCCATTCAGTATTCTACGAA...CCACATCAAAATGGGAAATACT 991
          |||:::   |||:::      |||   |||      |||:::      :
     442 eAspAsnValSerIleAsnAlaGluAspProLeuPheAsnAsnThrGlyI 459

     992 TTTTTAACGACAATAATAATGGCGCAGGAAAAATCGATGCCAAACATAAA 1041
         ::   ::::::      :::|||:::|||:::   :::          |||
     459 leAsnAspGluPheProThrGlyThrGlyGluAlaSerAspProLysLys 475

    1042 CACTATTCTCTACCTTATAGATTAAAAACACGAACCGTTCAATTGTTT.. 1089
         :::   :::|||         |||   :::   |||::::::   :::
     476 AsnSerGluLeuLysThrThrLeuThrAsnThrThrIleSerAsnTyrLe 492

    1090 ..................AATGTTTCTTTATCCGAGACAGCAAGAGAAC 1120
                            |||::::::   |||:::
     492 uLysAsnAlaTrpThrMetAsnIleThrAlaSerArgLysLeuThr.... 507

    1121 CTGTTTATCATGCTGCAGGTGGGGTCAACAGTTATCGACCCAGACTGAAT 1170
                                 |||||||||            :::|||
     508 ......................ValAsnSer.........SerIleAsn 513

    1171 AATGGAGAAAATATTTCCTTTATTGACAAAGGAAAAGGTGAATTGATACT 1220
           |||:::|||      :::|||      :::||||||
     514 IleGlySerAsnSerHisLeuIleLeuHisSerLysGly..... 526

    1221 TACCAGCAACATCAACCAAGGCGCGGGCGGTTTGTATTTTGAGGGTAATT 1270
                         |||   :::|||||||:::   ::::::|||::::
     527 ..............GlnArgGlyGlyGlyValGlnIleAspGlyAspI 538
```

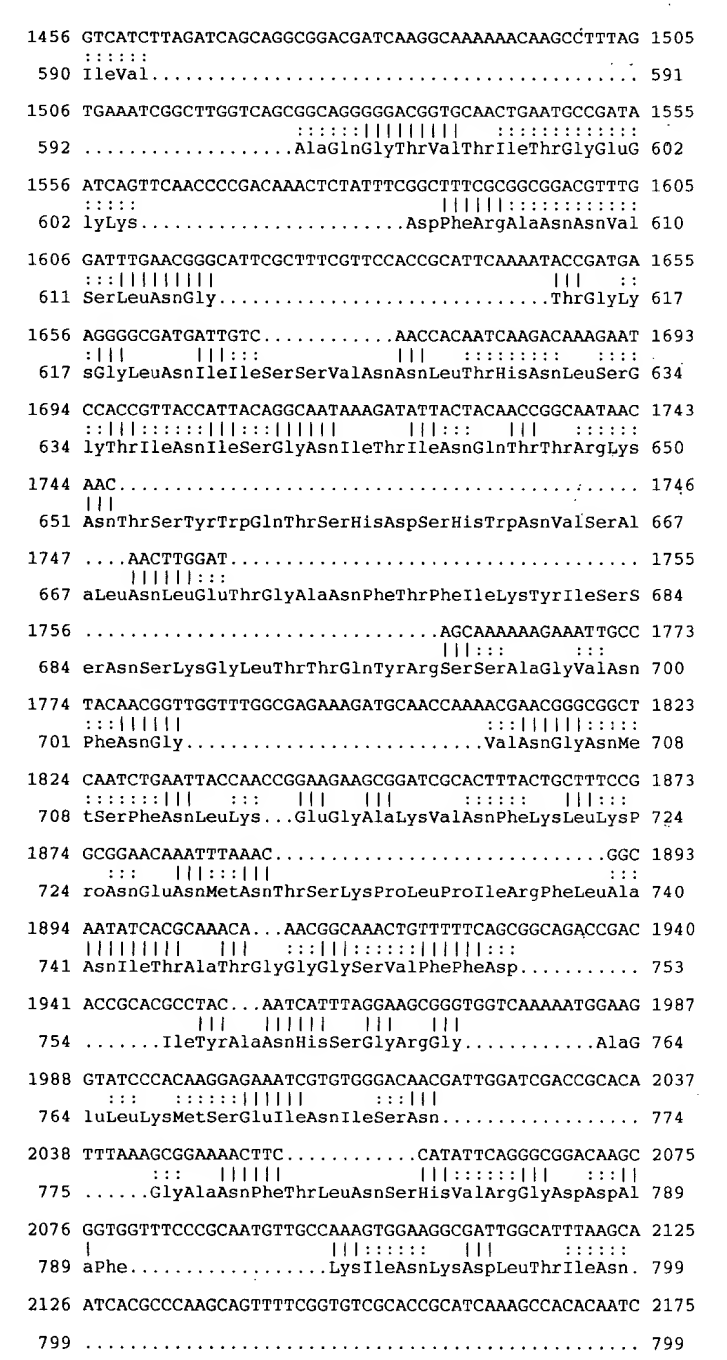


```
                 :::::::|||   :::   :::
  963 ..............IleGlyGlyAspIleIleAsn.............. 969

 2908 TTTATGTCGGAACTCTTCGGCTACCGCAGCGGCAAATTGAAGCTGGCGGA 2957
                              :::::::|||:::|||:::::::::::
  970 .........................LysLysGlySerLeuAsnIleThrAs 978

 2958 AAGTTCC.........................................G 2965
      :|||:::                                        |
  978 pSerAsnAsnAspAlaGluIleGlnIleGlyGlyAsnIleSerGlnLysG 995

 2966 AAGGCACTTACACCTTGGCT.........GTCAACAATACCGGCAACGAA 3006
      |||||:::   |||:::::::          :::|||   |||
  995 luGlyAsnLeuThrIleSerSerAspLysIleAsnIleThr......... 1008

 3007 CCCGTAAGTCTCGAGCAATTGACGGTA.........GTGGAAGGAAAAGA 3047
                   :::|||:::|||:::         ::::::|||:::||
 1009 ............LysGlnIleThrIleLysLysGlyIleAspGlyGluAs 1021

 3048 C...AACACACCGCTGTCCGAAAATCTTAATTTCACCCTGCAAAACGAAC 3094
      |   ::::::      ::::::|||   |||:::|||::::::
 1021 pSerSerSerAspAlaThrSerAsnAlaAsnLeuThrIleLys....... 1035

 3095 ACGTCGATGCCGGCGCATGGCGTTATCAGCTTATCCGCAAAGACGGCGAG 3144

 1035 ................................................... 1035

 3145 TTCCGCCTGCATAATCCGGTCAAAGAACAAGAGCTTTCCGACAAACTCGG 3194
                        :::|||||   :::|||:::::   |||::
 1036 ..................ThrLysGluLeuLysLeuThrGluAspLeuSe 1046

 3195 CAAGGCGGGAGAAACAGAGGCCGCCTTGACGGCAAAACAGGCACAACTTG 3244
      :   :::|||   :::::::|||   :::|||||||||:::::::::
 1046 rIleSerGlyPheAsnLysAlaGluIleThrAlaLysAspGlyArgAspL 1063

 3245 CCGCCAAACAACAGGCGGAAAAAGACAACGCGCAAAGC............ 3282
        :::      :::::::::      |||:::   :::
 1063 euThrIleGlyAsnSerAsnAspGlyAsnSerGlyAlaGluAlaLysThr 1079

 3283 ..................CTTGACGCGCTGATTGCGGCC...GGGCGCAA 3311
                         |||:::   |||:::|||   |||:::||
 1080 ValThrPheAsnAsnValLysAspSerLysIleSerAlaAspGlyHisAs 1096

 3312 TGCCACCGAAAAGGCAGAAAGTGTTGCCGAACCGGCCCGGCAGGCAGGCG 3361
      |:::|||         :::|||   ::::::   :::   :::   ||||
 1096 nValThr......LeuAsnSerLysValLysThrSerSerSerAsnGlyG 1111

 3362 GGGAAAATGCCGGCATTATGCAGGCGGAGGAAGAGAAAAAACGGGTGCAG 3411
      ||:::::::::                                    :::
 1111 lyArgGluSer.....................................Asn 1115

 3412 GCGGATAAAGACACCGCCTTG.........GCGAAACAGCGCGAAGCGGA 3452
      :::|||:::|||||||:::|||         ||||||:::   |||:::
 1116 SerAspAsnAspThrGlyLeuThrIleThrAlaLysAsnValGluVal.. 1131

 3453 AACCCGGCCGGCTACCACCGCCTTCCCCCGCGCCCGCCGCGCCCGCCGGG 3502

 1131 ................................................... 1131

 3503 ATTTGCCGCAACCGCAGCCCCAACCGCAACCCCAACCGCAGCGCGACCTG 3552
                                              :::::::|||:::
 1132 .....................................AsnLysAspIle 1135

 3553 ATCAGC...CGTTATGCCAATAGCGGTTTGAGTGAATTTTCCGCCACGCT 3599
        |||   :::   :::|||         ||||||      :::||| :
 1136 ThrSerLeuLysThrValAsnIleThrAlaSerGluLysValThrThr.T 1152

 3600 CAACAGCGTTTTCGCCGTACAGGACGAATTGGACCGCGTGTTTGCCGAAG 3649
      ::::::::   :::       ||||||      |||   :::
 1152 hrAlaGlySerThrIleAsnAlaThrAsnGlyLysAlaSerIleThrThr 1168

 3650 ACCGCCGCAACGCCGTTTGGACAAGCGGCATCCGGGACACCAAACACTAC 3699
        ::::::         |||      :::::::::|||   ::::::::
```

```
      1169 LysThrGlyAspIleSerGlyThrIleSerGlyAsnThrValSerValSe 1185

      3700 CGTTCGC...AAGATTTCCGCGCCTACCGCCAACAAACCGACCTGCGCCA 3746
           ::::       :::::::::   ::::::::::|||      :::   :
      1185 rAlaThrValAspLeuThrThrLysSerGlySerLysIleGluAlaLysS 1202

      3747 AATCGGTATGCAGAAAAACCTCGGCAGCGGGCGCGTCGGCATCCTGTTTT 3796
           :::::   ::::::   |||||||||:::|||:::   ::::::
      1202 erGlyGluAlaAsnValThrSerAlaThrGlyThrIleGlyGly..... 1216

      3797 CGCACAACCGGACCGGAAACACCTTCGACGACGGCATCGGCA........ 3838
              |||   :::   :::|||   ::::::||||||:::|||
      1217 ...ThrIleSerGlyAsnThrValAsnValThrAlaAsnAlaGlyAspLe 1232

      3839 ............................ACTCGGCACGGCTTGCCCACGG 3860
                                        :::   :::|||      |||:
      1232 uThrValGlyAsnGlyAlaGluIleAsnAlaThrGluGlyAlaAlaThrL 1249

      3861 TGCCGTTTTCGGGCAAT......ACGGCATCGGCAGGTTCGACATCGGCA 3904
           ::      :::|||||||      |||::::::||||||||:::   :::
      1249 euThrAlaThrGlyAsnThrLeuThrThrGluAlaGlySerSerIleThr 1265

      3905 TCAGCGCGGGCGCGGGTTTTAGTAGCGGCAGCCTTTCAGACGGCATCAGA 3954
           |||:::::::::::::|||   :::   |||      |||::::::|||
      1266 SerThrLysGlyGlnValAspLeuLeuAla......GlnAsnGlySerIl 1280

      3955 GGCAAAATCCGCCGCCGCGTGCTGCATTACGGCATTCAGGCAAGATACCG 4004
            |||   |||      |||:::   :::|||                |||
      1280 eAlaGlySerIleAsnAlaAlaAsnValThrLeuAsnThrThrGlyThrL 1297

      4005 CGCAGGTTTCGGCGGATTCGGCATCGAACCGCACATCGGCGCAACGCGCT 4054
               :::   |||   |||      :::   ||||||:::      :::
      1297 euThrThrValAlaGlySerAspIleLysAlaThrSerGlyThrLeuVal 1313

      4055 ATTTCGTCCAAA 4066
           |||:::::::|||
      1314 IleAsnAlaLys 1317
```

```
seq_name: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-08-719-641-4

seq_documentation_block:
; Sequence 4, Application US/08719641
; Patent No. 6218141
;  GENERAL INFORMATION:
;    APPLICANT:  Barenkamp, Stephen J
;    TITLE OF INVENTION:  High Molecular Weight Surface Proteins
;    TITLE OF INVENTION:  of No. 6218141-Typeable Haemophilus
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Shoemaker and Mattare, Ltd.
;      STREET:  2001 Jefferson Davis Hwy., 1203 Crystal Plaza
;      STREET:  Bldg. 1
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22202-0286
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/719,641
;      FILING DATE:
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/302,832
;      FILING DATE:  16-SEP-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US PCT/US93/02166
;      FILING DATE:  16-MAR-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9205704.1
```

```
;      FILING DATE:  16-MAR-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Berkstresser, Jerry W
;      REGISTRATION NUMBER:  22,651
;      REFERENCE/DOCKET NUMBER:  1038-625
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 415-0810
;      TELEFAX:  (703) 415-0813
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1477 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-719-641-4


alignment_scores:
            Quality:  274.50              Length:   1488
              Ratio:  0.419                Gaps:     73
 Percent Similarity:  44.019   Percent Identity: 19.220

alignment_block:
 US-09-303-518D-653 x US-08-719-641-4   ..

 Align seg 1/1  to: US-08-719-641-4  from: 1  to: 1477

     178 AATAAAGGCAAGTTTGCAGTCGGGGCGAAAGATATTGAGGTTTACAACAA 227
         |||   |||   ::::::||||||   ||||||   :::|||
     199 AsnHisGlyLeuIleThrValGly...LysAspGlySerVal........ 211

     228 AAAAGGGGAGTTGGTCGGCAAATCGATGACGAAAGCCCCGATGATTGATT 277
             :::|||:::|||   ::::::   :::
     212 .......AsnLeuIleGlyGlyLysValLysAsnGlu............. 221

     278 TTTCTGTGGTATCGCGTAACGGC......GTGGCGGCATTGGCGGGCGAT 321
           :::|||:::|||   ||||||      ::::::   |||||||||:::
     222 ..GlyValIleSerValAsnGlyGlySerIleSerLeuLeuAlaGlyGln 237

     322 CAA.................................TATATTGTGAG 335
         :::                                 |||   :::::
     238 LysIleThrIleSerAspIleIleAsnProThrIleThrTyrSerIleAl 254

     336 CGTGGCACATAACGGCGGCTATAACAATGTTGATTTTGGTGCGGAGGGAA 385
         ::::   :::|||   :::   |||      |||:::   |||:::|||:
     254 aAlaProGluAsnGluAlaValAsnLeuGlyAspIlePheAlaLysGlyG 271

     386 GCAAT.......................CCCGATCAGCACCGCTTTTCT 411
         ::|||                       :::|||   ::::::|||
     271 lyAsnIleAsnValArgAlaAlaThrIleArgAsnGlnGlyLysLeuSer 287

     412 TACCAAATTGTGAAAAGAAATAATTATAAAGCAGGGACTAACGGCCATCC 461
            :::   |||:::::::::      |||:::|||
     288 AlaAspSerValSerLysAsp......LysSerGly.............. 297

     462 TTATGGCGGCGATTATCATATGCCGCGTTTGCACAAATTTGTCACAGATG 511
                                         ::::::|||   :::|
     298 ................................AsnIleValLeuSerA 303

     512 CAGAACCTGTTGAG.................ATGACCAGTTATATGGAT 543
         ||:::      |||                 :::::::::      :::
     303 laLysGluGlyGluAlaGluIleGlyGlyValIleSerAlaGlnAsnGln 319

     544 GGGTGGAAATACGCTGATTTAAATAAATACCCTGATCGTGTT......CG 587
               |||   :::   |||           |||:::|||      ::
     320 GlnAlaLysGlyGlyLysLeuMetIleThrGlyAspLysValThrLeuLy 336

     588 AATCGGAGCA.........................GGCAGACAATATT 610
         :   ||||||                        |||:::   |||
     336 sThrGlyAlaValIleAspLeuSerGlyLysGluGlyGlyGluThrTyrL 353

     611 GGCGGTCTGATGAA...GACGAACCCAATAAC.................. 639
             :::|||||||      |||   :::|||

     353 euGlyGlyAspGluArgGlyGluGlyLysAsnGlyIleGlnLeuAlaLys 369

     640 ..............CGCGAAAGTTCATATCATATTGCAAGC........ 666
                         :::   |||:::   ::::::::::::
     370 LysThrSerLeuGluLysGlySerThrIleAsnValSerGlyLysGluLy 386

     667 ..........GCATATTCTTGG............CTCGTCGGTGGCAATA 694
                    |||      |||            |||:::   ||||||
     386 sGlyGlyPheAlaIleValTrpGlyAspIleAlaLeuIleAspGlyAsnI 403

     695 CCTTTGCACAAAATGGATCAGGTGGTGGCACAGTCAACTTAGGTAGCGAA 744
              ||||||   |||||||||
     403 leAsnAlaGln...GlySerGly.............................. 409

     745 AAAATTAAACATAGCCCATATGGTTTTTTACCAACAGGAGGCTCATTTGG 794
                                    :::      |||||||||
     410 .....................AspIleAlaLysThrGlyGlyPheValGl 419

     795 CGACAGTGGCTCACCAATGTTTATCTATGATGCCCAAAAGCAAAAGTGGT 844
             ||||||      :::|||||||   |||
     419 uThrSerGlyHisAspLeuPheIleLysAsp.................... 429

     845 TAATTAATGGGGTATTGCAAACAGGCAACCCCTATATAGGAAAAAGCAAT 894
              |||:::::::::::::::
     430 .....AsnAlaIleValAspAla............................ 435

     895 GGCTTCCAGCTAGTTCGTAAAGATTGGTTCTATGATGAAATCTTTGCTGG 944
                           |||:::|||:::   |||
     436 ..................LysGluTrpLeuLeuAsp............Ph 442

     945 AGATACCCATTCAGTATTCTACGAA...CCACATCAAAATGGGAAATACT 991
         |||:::   |||:::      |||   |||      |||:::      :
     442 eAspAsnValSerIleAsnAlaGluAspProLeuPheAsnAsnThrGlyI 459

     992 TTTTTAACGACAATAATAATGGCGCAGGAAAAATCGATGCCAAACATAAA 1041
         ::   ::::::      :::|||:::|||:::   :::          |||
     459 leAsnAspGluPheProThrGlyThrGlyGluAlaSerAspProLysLys 475

    1042 CACTATTCTCTACCTTATAGATTAAAAACACGAACCGTTCAATTGTTT.. 1089
         :::   :::|||         |||   :::   |||::::::   :::
     476 AsnSerGluLeuLysThrThrLeuThrAsnThrThrIleSerAsnTyrLe 492

    1090 ..................AATGTTTCTTTATCCGAGACAGCAAGAGAAC 1120
                            |||:::::   |||:::
     492 uLysAsnAlaTrpThrMetAsnIleThrAlaSerArgLysLeuThr.... 507

    1121 CTGTTTATCATGCTGCAGGTGGGGTCAACAGTTATCGACCCAGACTGAAT 1170
                              |||||||||           :::|||
     508 ......................ValAsnSer.........SerIleAsn 513

    1171 AATGGAGAAAATATTTCCTTTATTGACAAAGGAAAAGGTGAATTGATACT 1220
            |||:::|||      :::|||      :::||||||
     514 IleGlySerAsnSerHisLeuIleLeuHisSerLysGly........... 526

    1221 TACCAGCAACATCAACCAAGGCGCGGGCGGTTTGTATTTTGAGGGTAATT 1270
                         |||   :::||||||:::   :::::::|||::::
     527 ..............GlnArgGlyGlyGlyValGlnIleAspGlyAspI 538

    1271 TTACGGTCTCGCCTAAAAACAACGAAACGTGGCAAGGCGCGGGCGTTCAT 1320
         ::|||   :::      |||         ::::::|||:::   |||
     538 leThrSerLysGlyGlyAsnLeuThrIleTyrSerGlyGlyTrpValAsp 554

    1321 ATCAGTGATGGCAGTACCGTTACTTGGAAAGTAAACGGCGTGGCAAACGA 1370
         :::      :::   |||
     555 ValHisLysAsnIleThr.............................. 560

    1371 CCGCCTGTCCAAAATCGGCAAAGGCACGCTGCTGGTTCAAGCC....... 1413
                    :::   :::|||   |||   :::   |||
     561 .............LeuAspGlnGlyPheLeuAsnIleThrAlaAlaSerV 573

    1414 ........AAAGGGGAAAACCAAGGCTCGGTCAGCGTGGGCGACGGTAAA 1455
                 :::|||   |||:::   :::   ::::::::::::::|||
     573 alAlaPheGluGlyGlyAsnAsnLysAlaArgAspAlaAlaAsnAlaLys 589
```


1456 GTCATCTTAGATCAGCAGGCGGACGATCAAGGCAAAAAACAAGCCTTTAG 1505
590 IleVal 591
1506 TGAAATCGGCTTGGTCAGCGGCAGGGGGACGGTGCAACTGAATGCCGATA 1555
592 AlaGlnGlyThrValThrIleThrGlyGluG 602
1556 ATCAGTTCAACCCCGACAAACTCTATTTCGGCTTTCGCGGCGGACGTTTG 1605
602 lyLys AspPheArgAlaAsnAsnVal 610
1606 GATTTGAACGGGCATTCGCTTTCGTTCCACCGCATTCAAAATACCGATGA 1655
611 SerLeuAsnGly ThrGlyLy 617
1656 AGGGGCGATGATTGTC AACCACAATCAAGACAAAGAAT 1693
617 sGlyLeuAsnIleIleSerSerValAsnAsnLeuThrHisAsnLeuSerG 634
1694 CCACCGTTACCATTACAGGCAATAAAGATATTACTACAACCGGCAATAAC 1743
634 lyThrIleAsnIleSerGlyAsnIleThrIleAsnGlnThrThrArgLys 650
1744 AAC 1746
651 AsnThrSerTyrTrpGlnThrSerHisAspSerHisTrpAsnValSerAl 667
1747 ....AACTTGGAT 1755
667 aLeuAsnLeuGluThrGlyAlaAsnPheThrPheIleLysTyrIleSerS 684
1756 AGCAAAAAAGAAATTGCC 1773
684 erAsnSerLysGlyLeuThrThrGlnTyrArgSerSerAlaGlyValAsn 700
1774 TACAACGGTTGGTTTGGCGAGAAAGATGCAACCAAAACGAACGGGCGGCT 1823
701 PheAsnGly ValAsnGlyAsnMe 708
1824 CAATCTGAATTACCAACCGGAAGAAGCGGATCGCACTTTACTGCTTTCCG 1873
708 tSerPheAsnLeuLys...GluGlyAlaLysValAsnPheLysLeuLysP 724
1874 GCGGAACAAATTTAAAC GGC 1893
724 roAsnGluAsnMetAsnThrSerLysProLeuProIleArgPheLeuAla 740
1894 AATATCACGCAAACA...AACGGCAAACTGTTTTTCAGCGGCAGACCGAC 1940
741 AsnIleThrAlaThrGlyGlyGlySerValPhePheAsp 753
1941 ACCGCACGCCTAC...AATCATTTAGGAAGCGGGTGGTCAAAAATGGAAG 1987
754 IleTyrAlaAsnHisSerGlyArgGly AlaG 764
1988 GTATCCCACAAGGAGAAATCGTGTGGGACAACGATTGGATCGACCGCACA 2037
764 luLeuLysMetSerGluIleAsnIleSerAsn 774
2038 TTTAAAGCGGAAAACTTC CATATTCAGGGCGGACAAGC 2075
775 GlyAlaAsnPheThrLeuAsnSerHisValArgGlyAspAspAl 789
2076 GGTGGTTTCCCGCAATGTTGCCAAAGTGGAAGGCGATTGGCATTTAAGCA 2125
789 aPhe LysIleAsnLysAspLeuThrIleAsn 799
2126 ATCACGCCCAAGCAGTTTTCGGTGTCGCACCGCATCAAAGCCACACAATC 2175
799 799
2176 TGTACACGTTCGGACTGGACGGGTCTGACAAGTTGTACCGAAAAAACCAT 2225
800 AlaThrAsnSerAsnPhe 805
2226 TACCGACGATAAAGTGATTGCTTCATTGAGCAAGACCGACATCAGAGGCA 2275
806 SerLeuArgGlnThrLysAspAspPheT 815
2276 ATGTCAGCCTTGCCGATCACGCTCATTTAAATCTCACAGGACTTGCCACA 2325
815 yrAspGlyTyrAlaArgAsnAlaIleAsnSerThrTyrAsnIleSerIle 831
2326 CTCAACGGCAATCTTAGTGCAGGCGGAGACACGCACTATACGGTTACGCG 2375
832 LeuGlyGlyAsnValThrLeuGlyGly 840
2376 CAACGCCACCCAAAACGGCAACCTCAGCCTCGTGGGCAATGCCCAAGCAA 2425
841 GlnAsnSerSerSerSerIleThrGlyAsn IleT 852
2426 CATTTAATCAAGCC ACATTAAACGGCAACACATCGGCTTCG 2466
852 hrIleGluLysAlaAlaAsnValThrLeuGluAlaAsnAsnAlaProAsn 868
2467 GACAATGCTTCATTTAATCTAAGCAACAACGCCGTACAAAACGGCAGTCT 2516
869 GlnGln AsnIleArgAspArgValIleLysLeuGlySerLe 882
2517 GACGCTTTCCGACAACGCTAAGGCAAACGTAAGCCATTCCGCACTCAACG 2566
882 uLeuValAsnGlySerLeuSerLeuThrGlyGluAsnAlaAspIleLysG 899
2567 GCAATGTCTCCCTAGCCGATAAGGCAGTATTCCAT TTTGAA 2607
899 lyAsnLeuThrIleSerGluSerAlaThrPheLysGlyLysThrArgAsp 915
2608 AACAGCCGCTTTACCGGAAAAATCAGCGGCGGCAAGGATACGGCATTACA 2657
916 ThrLeuAsnIleThrGlyAsnPheThrAsnAsnGlyThrAlaGluIleAs 932
2658 CTTAAAAGACAGCGAATGGACGCTGCCGTCGGGCACGGAATTAGGCAATT 2707
932 nIleThrGlnGlyValValLysLeuGlyAsnValThrAsnAspGlyAspL 949
2708 TAAACCTTGACAACGCCACCATTACACTCAATTCCGCCTATCGACACGAT 2757
949 euAsn IleThrThrHisAla 955
2758 GCGGCAGGCGCGCAAACCGGCAGTGCGGCAGATGCGCCGCGCCGCCGTTC 2807
956 LysArgAs 958
2808 GCGCCGTTCCCTATTATCCGTTACGCCGCCAACTTCGGCAGAATCCCGTT 2857
958 nGlnArgSerIle 962
2858 TCAACACGCTGACGGTAAACGGCAAATTGAACGGTCAGGGAACATTCCGC 2907
963 IleGlyGlyAspIleIleAsn 969
2908 TTTATGTCGGAACTCTTCGGCTACCGCAGCGGCAAATTGAAGCTGGCGGA 2957
970 LysLysGlySerLeuAsnIleThrAs 978
2958 AAGTTCC G 2965
978 pSerAsnAsnAspAlaGluIleGlnIleGlyGlyAsnIleSerGlnLysG 995
2966 AAGGCACTTACACCTTGGCT GTCAACAATACCGGCAACGAA 3006
995 luGlyAsnLeuThrIleSerSerAspLysIleAsnIleThr 1008
3007 CCCGTAAGTCTCGAGCAATTGACGGTA GTGGAAGGAAAAGA 3047

```
                 :::|||:::|||:::          ::::::|||:::||
1009 ............LysGlnIleThrIleLysLysGlyIleAspGlyGluAs 1021

3048 C...AACACACCGCTGTCCGAAAATCTTAATTTCACCCTGCAAAACGAAC 3094
     |   ::::::      ::::::|||   |||:::|||::::::
1021 pSerSerSerAspAlaThrSerAsnAlaAsnLeuThrIleLys....... 1035

3095 ACGTCGATGCCGGCGCATGGCGTTATCAGCTTATCCGCAAAGACGGCGAG 3144

1035 ..................................................... 1035

3145 TTCCGCCTGCATAATCCGGTCAAAGAACAAGAGCTTTCCGACAAACTCGG 3194
                      :::||||||   :::|||::::::   |||::
1036 ..................ThrLysGluLeuLysLeuThrGluAspLeuSe 1046

3195 CAAGGCGGGAGAAACAGAGGCCGCCTTGACGGCAAAACAGGCACAACTTG 3244
     :   :::|||   :::::::|||   :::|||||||||:::::::::
1046 rIleSerGlyPheAsnLysAlaGluIleThrAlaLysAspGlyArgAspL 1063

3245 CCGCCAAACAACAGGCGGAAAAAGACAACGCGCAAAGC............. 3282
       :::      :::::::::      |||:::   :::
1063 euThrIleGlyAsnSerAsnAspGlyAsnSerGlyAlaGluAlaLysThr 1079

3283 ..................CTTGACGCGCTGATTGCGGCC...GGGCGCAA 3311
                         |||:::   |||:::|||   |||:::||
1080 ValThrPheAsnAsnValLysAspSerLysIleSerAlaAspGlyHisAs 1096

3312 TGCCACCGAAAAGGCAGAAAGTGTTGCCGAACCGGCCCGGCAGGCAGGCG 3361
     |:::|||         :::|||   ::::::   :::   :::   ||||
1096 nValThr......LeuAsnSerLysValLysThrSerSerSerAsnGlyG 1111

3362 GGGAAAATGCCGGCATTATGCAGGCGGAGGAAGAGAAAAAACGGGTGCAG 3411
     ||:::::::::                                  :::
1111 lyArgGluSer.................................Asn 1115

3412 GCGGATAAAGACACCGCCTTG.........GCGAAACAGCGCGAAGCGGA 3452
     :::|||:::||||||:::|||         ||||||:::   |||:::
1116 SerAspAsnAspThrGlyLeuThrIleThrAlaLysAsnValGluVal.. 1131

3453 AACCCGGCCGGCTACCACCGCCTTCCCCCGCGCCCGCCGCGCCCGCCGGG 3502

1131 ..................................................... 1131

3503 ATTTGCCGCAACCGCAGCCCCAACCGCAACCCCAACCGCAGCGCGACCTG 3552
                                          :::::::|||:::
1132 .....................................AsnLysAspIle 1135

3553 ATCAGC...CGTTATGCCAATAGCGGTTTGAGTGAATTTTCCGCCACGCT 3599
        |||   :::   :::|||         ||||||      :::||| :
1136 ThrSerLeuLysThrValAsnIleThrAlaSerGluLysValThrThr.T 1152

3600 CAACAGCGTTTTCGCCGTACAGGACGAATTGGACCGCGTGTTTGCCGAAG 3649
     ::::::::   :::         ||||||      |||   ::::
1152 hrAlaGlySerThrIleAsnAlaThrAsnGlyLysAlaSerIleThrThr 1168

3650 ACCGCCGCAACGCCGTTTGGACAAGCGGCATCCGGGACACCAAACACTAC 3699
        ::::::         |||      ::::::::::|||   ::::::::
1169 LysThrGlyAspIleSerGlyThrIleSerGlyAsnThrValSerValSe 1185

3700 CGTTCGC...AAGATTTCCGCGCCTACCGCCAACAAACCGACCTGCGCCA 3746
     ::::        :::::::::   ::::::::::|||      :::   :
1185 rAlaThrValAspLeuThrThrLysSerGlySerLysIleGluAlaLysS 1202

3747 AATCGGTATGCAGAAAAACCTCGGCAGCGGGCGCGTCGGCATCCTGTTTT 3796
     :::::   ::::::   |||||||||:::|||:::   ::::::
1202 erGlyGluAlaAsnValThrSerAlaThrGlyThrIleGlyGly...... 1216

3797 CGCACAACCGGACCGGAAACACCTTCGACGACGGCATCGGCA........ 3838
        |||   :::   :::|||   ::::::||||||:::|||
1217 ...ThrIleSerGlyAsnThrValAsnValThrAlaAsnAlaGlyAspLe 1232

3839 ........................ACTCGGCACGGCTTGCCCACGG 3860
                               :::   :::|||      |||:
```

```
     1232 uThrValGlyAsnGlyAlaGluIleAsnAlaThrGluGlyAlaAlaThrL 1249

     3861 TGCCGTTTTCGGGCAAT......ACGGCATCGGCAGGTTCGACATCGGCA 3904
          ::      :::||||||      |||::::::||||||||:::   :::
     1249 euThrAlaThrGlyAsnThrLeuThrThrGluAlaGlySerSerIleThr 1265

     3905 TCAGCGCGGGCGCGGGTTTTAGTAGCGGCAGCCTTTCAGACGGCATCAGA 3954
          |||:::::::::::|||   :::   |||      |||::::::|||
     1266 SerThrLysGlyGlnValAspLeuLeuAla......GlnAsnGlySerIl 1280

     3955 GGCAAAATCCGCCGCCGCGTGCTGCATTACGGCATTCAGGCAAGATACCG 4004
           |||   |||      |||:::   :::|||              |||
     1280 eAlaGlySerIleAsnAlaAlaAsnValThrLeuAsnThrThrGlyThrL 1297

     4005 CGCAGGTTTCGGCGGATTCGGCATCGAACCGCACATCGGCGCAACGCGCT 4054
               :::   |||   |||      :::   ||||||:::      :::
     1297 euThrThrValAlaGlySerAspIleLysAlaThrSerGlyThrLeuVal 1313

     4055 ATTTCGTCCAAA 4066
          |||:::::::|||
     1314 IleAsnAlaLys 1317
```

```
seq_name: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-08-460-269C-4

seq_documentation_block:
; Sequence 4, Application US/08460269C
; Patent No. 6197548
;   GENERAL INFORMATION:
;        APPLICANT: CLARE, JEFFREY J.
;                   ROMANOS, MICHAEL A.
;        TITLE OF INVENTION: EXPRESSION OF HETEROLOGOUS PROTEIN IN
;                            YEAST
;        NUMBER OF SEQUENCES: 17
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Millen, White, Zelano & Branigan, P.C.
;             STREET: 2200 Clarendon Blvd., Suite 1400
;             CITY: ARLINGTON
;             STATE: VA
;             COUNTRY: USA
;             ZIP: 22201
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/08/460,269C
;             FILING DATE: 02-Jun-1995
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Lebovitz, Richard M.
;             REGISTRATION NUMBER: 37,067
;             REFERENCE/DOCKET NUMBER: Popov-2
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (703) 243-6333
;             TELEFAX: (703) 243-6410
;   INFORMATION FOR SEQ ID NO: 4:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 911 amino acids
;             TYPE: amino acid
;             TOPOLOGY: linear
;        MOLECULE TYPE: protein
;        SEQUENCE DESCRIPTION: SEQ ID NO: 4:
US-08-460-269C-4


alignment_scores:
           Quality:  274.00            Length:   1149
             Ratio:  0.584               Gaps:     50
 Percent Similarity:  40.818   Percent Identity: 20.540

alignment_block:
 US-09-303-518D-653 x US-08-460-269C-4   ..
```

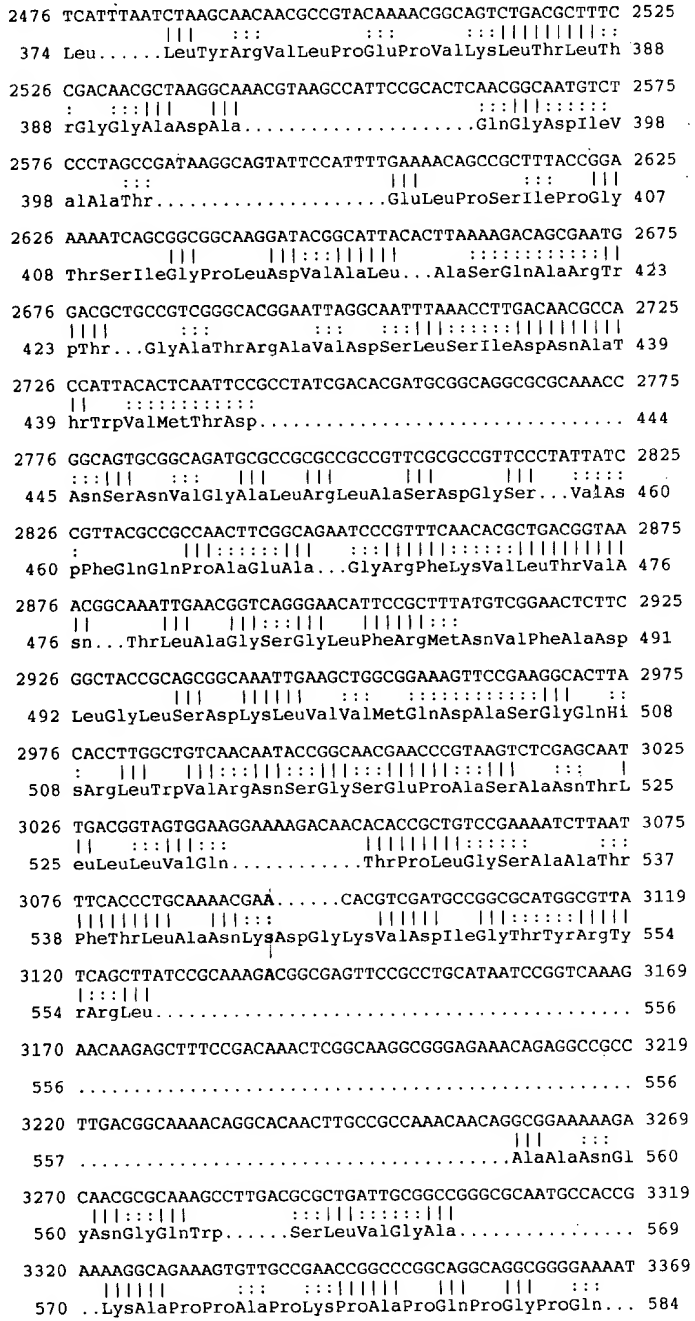
Align seg 1/1 to: US-08-460-269C-4 from: 1 to: 911

```
2947 AAGCTGGCGGAAAGTTCCGAAGGCACTTACACCTTGGCTGTCAACAATAC 2996
490 ValValMetArgAspAlaSerGlyGlnHisArgLeuLeuValArgAsnSe 506

2997 CGGCAACGAACCCGTAAGTCTCGAGCAATTGACGGTAGTGGAAGGAAAAG 3046
506 rGlySerGluProAlaSerGlyAsnThrMetLeuLeuValGln..... 520

3047 ACAACACACCGCTGTCCGAAAATCTTAATTTCACCCTGCAAAACGAA... 3093
521 .....ThrProArgGlySerAlaAlaThrPheThrLeuAlaAsnLysAsp 535

3094 ...CACGTCGATGCCGGCGCATGGCGTTATCAGCTTATCCGCAAA...GA 3137
536 GlyLysValAspIleGlyThrTyrArgTyrArgLeuAlaAlaAsnGlyAs 552

3138 CGGCGAGTTCCGCCTGCATAATCCGGTCAAAGAACAAGAGCTTTCCGACA 3187
552 nGlyGlnTrpSerLeu.Val..... 558

3188 AACTCGGCAAGGCGGGAGAAACAGAGGCCGCCTTGACGGCAAAACAGGCA 3237
558 ..... 558

3238 CAACTTGCCGCCAAACAACAGGCGGAAAAAGACAACGCGCAAAGCCTTGA 3287
558 ..... 558

3288 CGCGCTGATTGCGGCCGGGCGCAATGCCACCGAAAAGGCAGAAAGTGTTG 3337
559 .....GlyAlaLysAlaProProAlaProLysProAla 569

3338 CCGAACCGGCCCGGCAGGCAGGCGGGGAAAATGCCGGCATTATGCAGGCG 3387
570 Pro...GlnProGlyProGln..... 575

3388 GAGGAAGAGAAAAAACGGGTGCAGGCGGATAAAGACACCGCCTTGGCGAA 3437
575 ..... 575

3438 ACAGCGCGAAGCGGAAACCCGGCCGGCTACCACCGCCTTCCCCCGCGCCC 3487
576 .....ProGlyProGlnProProGln.ProProGlnPr 586

3488 GCCGCGCCCGCCGGGATTTGCCGCAACCGCAGCCCCAACCGCAACCCCAA 3537
586 oProGlnProProGlnArgGlnProGluAla.ProAlaProGlnProPro 602

3538 CCGCAGCGCGACCTGATCAGCCGTTATGCCAATAGCGGTTTGAGTGAATT 3587
603 AlaGlyArgGluLeu...SerAlaAlaAlaAsnAlaAlaValAsnThrGl 618

3588 TTCCGCCACGCTCAACAGCGTT......TTCGCCGTACAGGACGAATTGG 3631
618 yGlyValGlyLeuAlaSerThrLeuTrpTyrAlaGluSerAsnAlaLeuS 635

3632 ACCGCGTGTTTGCCGAAGACCGCCGCAAC.....GCCGTTTGG 3669
635 erLysArgLeuGlyGluLeuArgLeuAsnProAspAlaGlyGlyAlaTrp 651

3670 ACAAGCGGCATCCGGGACACCAAACACTACCGTTCGCAAGATTTCCGCGC 3719
652 GlyArgGlyPhe.....Al 656

3720 CTACCGCCAACAAACCGACCTGCGCCAAATCGGTATGCAGAAAAACCTCG 3769
656 aGlnArgGlnGlnLeuAsp..... 662

3770 GCAGCGGGCGCGTCGGCATCCTGTTTTCGCACAACCGGACCGGAAACACC 3819
663 .....AsnArgAlaGlyArgArg 668

3820 TTCGACGAC.....GGCATCGGCAACTCGGCACG 3848
```

```
669 PheAspGlnLysValAlaGlyPheGluLeuGlyAlaAspHisAlaValAl 685

3849 GCTTGCCCACGGTGCC......GTTTTCGGGCAATACGGCATCGGCAGGT 3892
685 aValAlaGlyGlyArgTrpHisLeuGlyGlyLeuAlaGlyTyrThrArgG 702

3893 TCGACATCGGCATCAGCGCGGGCGCGGGTTTTAGTAGCGGCAGCCTT... 3939
702 lyAspArgGlyPheThrGlyAspGlyGlyGlyHisThrAspSerValHis 718

3939 ..... 3939

719 ValGlyGlyTyrAlaThrTyrIleAlaAsnSerGlyPheTyrLeuAspAl 735

3940 .....TCAG 3943
735 aThrLeuArgAlaSerArgLeuGluAsnAspPheLysValAlaGlySerA 752

3944 ACGGC......ATCAGAGGCAAAATCCGCCGCCGCGTGCTGCATTACGGC 3987
752 spGlyTyrAlaValLysGlyLysTyrArgThr.....HisGly 764

3988 ATTCAGGCAAGATACCGCGCAGGT.....TTCGGCGGATT 4022
765 ValGlyAlaSerLeuGluAlaGlyArgArgPheAlaHisAlaAspGlyTr 781

4023 CGGCATCGAACCGCACATCGGCGCAACGCGCTATTTCGTCCAAAAAGCGG 4072
781 pPheLeuGluProGlnAlaGluLeuAlaValPheArgValGlyGlyGlyS 798

4073 ATTACCGATACGAAAAC...GTCAATATCGCCACCCCGGGCCTTGCATTC 4119
798 erTyrArgAlaAlaAsnGlyLeuArgValArgAspGluGlyGlySerSer 814

4120 AACCGCTACCGCGCGGGCATTAAGGCAGATTATTCATTCAAACCGGCGCA 4169
815 ValLeuGlyArgLeuGlyLeuGluValGlyLysArgIleGluLeuAlaGl 831

4170 ACACATTTCCATCACGCCTTATTTGAGCCTGTCCTATACCGATGCCGCTT 4219
831 yGlyArgGlnValGlnProTyrIleLysAlaSerValLeuGlnGluPheA 848

4220 CCGGCAAAGTCCGAACGCGCGTCAATACCGCCGTATTGGCGCAGGATTTC 4269
848 spGlyAlaGlyThrValArgThrAsnGlyIleAlaHisArgThrGluLeu 864

4270 GGCAAAACCCGCAGTGCGGAATGGGGCGTAAACGCCGAAATCAAAGGTTT 4319
865 ArgGlyThrArg...AlaGluLeuGly..... 872

4320 CACGCTGTCCCTCCACGCTGCCGCCGCCAAGGGGCCGCAATTGGAAGCG. 4368
873 ....LeuGlyMetAlaAlaAlaLeuGlyArgGlyHisSerLeuTyrAlaS 888

4369 .....CAGCACAGCGCGGGCATCAAATTAGGCTACCGCTGG 4404
888 erTyrGluTyrSerLysGlyProLysLeuAlaMetProTrp 901
```

```
seq_name: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-08-460-269C-2

seq_documentation_block:
; Sequence 2, Application US/08460269C
; Patent No. 6197548
;   GENERAL INFORMATION:
;        APPLICANT: CLARE, JEFFREY J.
;                   ROMANOS, MICHAEL A.
;        TITLE OF INVENTION: EXPRESSION OF HETEROLOGOUS PROTEIN IN
;                            YEAST
;        NUMBER OF SEQUENCES: 17
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Millen, White, Zelano & Branigan, P.C.
;             STREET: 2200 Clarendon Blvd., Suite 1400
```

```
;             CITY: ARLINGTON
;             STATE: VA
;             COUNTRY: USA
;             ZIP: 22201
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/08/460,269C
;             FILING DATE: 02-Jun-1995
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Lebovitz, Richard M.
;             REGISTRATION NUMBER: 37,067
;             REFERENCE/DOCKET NUMBER: Popov-2
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (703) 243-6333
;             TELEFAX: (703) 243-6410
;    INFORMATION FOR SEQ ID NO: 2:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 910 amino acids
;             TYPE: amino acid
;             TOPOLOGY: linear
;        MOLECULE TYPE: protein
;        SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-08-460-269C-2


alignment_scores:
            Quality:  268.00            Length:   1156
              Ratio:  0.551               Gaps:     53
 Percent Similarity:  42.042   Percent Identity: 21.367

alignment_block:
 US-09-303-518D-653 x US-08-460-269C-2   ..

 Align seg 1/1  to: US-08-460-269C-2  from: 1  to: 910

    1366 AACGACCGCCTGTCCAAAATCGGCAAAGGC..............ACGCT 1400
         |||      ||||||:::|||   |||:::               |||
       2 AsnMetSerLeuSerArgIleValLysAlaAlaProLeuArgArgThrTh 18

    1401 GCTGGTTCAAGCCAAAGGG............................. 1419
          |||::::::|||   |||
      18 rLeuAlaMetAlaLeuGlyAlaLeuGlyAlaAlaProAlaAlaHisAlaA 35

    1420 .....GAAAACCAAGGCTCGGTCAGCGTGGGCGACGGTAAAGTCATCTTA 1464
          :::||||||:::   |||::::::|||:::   :::      :::
      35 spTrpAsnAsnGlnSerIleValLysThrGlyGluArgGlnHisGlyIle 51

    1465 GATCAGCAGGCGGACGATCAAGGCAAAAAACAAGCCTTTAGTGAAATCGG 1514
            |||:::::::|||   |||      ::::::   |||
      52 HisIleGlnGlySerAspProGlyGlyValArgThrAlaSerGlyThrTh 68

    1515 CTTG...GTCAGCGGCAGGGGGACGGTGCAACTGAATGCCGATAATCAGT 1561
          :::   ||||||||||||   :::      :::      :::|||
      68 rIleLysValSerGlyArgGlnAlaGlnGlyIleLeuLeuGluAsn.... 83

    1562 TCAACCCCGACAAACTCTATTTCGGCTTTCGCGGCGGACGTTTGGATTTG 1611
                          :::   ||||||:::|||   :::
      84 ........ProAlaAlaGluLeuGlnPheArgAsnGlySerValThrSer 97

    1612 AACGGGCATTCGCTTTCGTTCCACCGCATTCAAAATACCGATGAAGGGGC 1661
         :::|||                        |||   :::|||:::|||
      98 SerGly.......................GlnLeuSerAspAspGlyIl 106

    1662 GATGATTGTCAACCACAATCAAGACAAAGAATCCACCGTTACCATTACAG 1711
                           :::         :::||||||||||:::   :
     106 e.................ArgArgPheLeuGlyThrValThrValLysA 117

    1712 GCAATAAA.........GATATTACTACAACCGGCAATAACAACAACTTG 1752
         :::::|||         |||   :::|||   :::|||   ::::::

     117 laGlyLysLeuValAlaAspHisAlaThrLeuAlaAsnValGlyAspThr 133

    1753 GATAGCAAAAAAGAAATTGCCTACAACGGTTGGTTTGGCGAGAAAGATGC 1802
                                          |||      :::      ||
     134 ............................TrpAspAspAspGlyIleAl 140

    1803 AACCAAAACGAACGGGCGGCTCAATCTGAATTACCAACCGGAAGAAGCGG 1852
         |      :::   |||:::         :::   |||   :::   ||||
     140 aLeuTyrValAlaGlyGlu.........GlnAlaGlnAlaSerIleAlaA 154

    1853 ATCGCACTTTACTGCTTTCCGGCGGAACAAATTTA......AACGGCAAT 1896
         ||   ||||||      :::||||||:::::::::      ::::::|||
     154 spSerThrLeuGlnGlyAlaGlyGlyValGlnIleGluArgGlyAlaAsn 170

    1897 ATCACGCAAACAAACGGCAAACTGTTTTTCAGCGGCAGACCGACACCGCA 1946
         :::|||      ::::::   :::      :::|||
     171 ValThrValGlnArgSerAlaIleValAspGlyGlyLeu........... 183

    1947 CGCCTACAATCATTTAGGAAGCGGGTGGTCA...AAAATGGAAGGTATCC 1993
                   |||:::|||:::      |||   :::   |||   :::|
     184 .........HisIleGlyAlaLeuGlnSerLeuGlnProGluAspLeuP 197

    1994 CACAAGGAGAAATCGTGTGGGACAACGATTGGATCGACCGCACA...... 2037
         ||   ::::::::::|||      :::      :::      :::
     197 roProSerArgValValLeuArgAspThrAsnValThrAlaValProAla 213

    2038 .....................................TTTAAAGC 2045
                                              :::
     214 SerGlyAlaProAlaAlaValSerValLeuGlyAlaSerGluLeuThrLe 230

    2046 GGAAAACTTCCATATTCAGGGCGGACAAGCGGTGGTTTCCCGCAATGTTG 2095
          ::::::   ||||||   ||||||:::|||:::         :::||||
     230 uAspGlyGlyHisIleThrGlyGlyArgAlaAla.........GlyValA 244

    2096 CCAAAGTGGAAGGCGATTGGCATTTAAGCAATCACGCCCAAGCAGTTTTC 2145
         ||                                          |||:::
     244 la.....................................AlaMetGln 247

    2146 GGTGTCGCACCGCATCAAAGCCACACAATCTGTACACGTTCGGAC..... 2190
         |||::::::   |||   :::::::::         |||:::|||
     248 GlyAlaValValHisLeuGlnArgAlaThrIleArgArgGlyAspAlaLe 264

    2190 ............................................... 2190

     264 uAlaGlyGlyAlaValProGlyGlyAlaValProGlyGlyAlaValProG 281

    2191 .....................................TGGACGGGT 2199
                                              |||   |||
     281 lyGlyPheGlyProGlyGlyPheGlyProValLeuAspGlyTrpTyrGly 297

    2200 CTGACAAGTTGTACCGAAAAAACCATTACCGACGATAAAGTGATTGCTTC 2249
         :::
     298 Val........................................... 298

    2250 ATTGAGCAAGACCGACATCAGAGGC...AATGTCAGCCTTGCCGATCACG 2296
                       |||:::   |||   :::|||:::||||||:::
     299 .............AspValSerGlySerSerValGluLeuAlaGln.... 309

    2297 CTCATTTAAATCTCACAGGACTTGCCACACTCAACGGCAATCTTAGTGCA 2346
                ::::::::            |||::::::   :::   :::
     310 ........SerIleValGluAlaProGluLeuGlyAlaAlaIleArgVal 323

    2347 GGCGGAGACACGCACTATACGGTTACGCGCAAC......GCCACCCAAAA 2390
         |||      ::::::   ||||||      :::      ::::::   ::
     324 GlyArgGlyAlaArgValThrValProGlyGlySerLeuSerAlaProHi 340

    2391 CGGCAACCTCAGCCTCGTGGGCAATGCCCAAGCAACATTTAATCAAGCC. 2439
         :|||||:::      :::|||:::|||:::            ||||||
     340 sGlyAsnValIleGluThrGlyGlyAlaArgArgPheAlaProGlnAlaA 357

    2440 .............ACATTAAACGGCAACACATCGGCTTCGGACAATGCT 2475
                       ||||||:::::::::::::   |||:::   :::|||
     357 laProLeuSerIleThrLeuGlnAlaGlyAlaHisAlaGlnGlyLysAla 373
```


2476 TCATTTAATCTAAGCAACAACGCCGTACAAAACGGCAGTCTGACGCTTTC 2525
374 Leu.....LeuTyrArgValLeuProGluProValLysLeuThrLeuTh 388
2526 CGACAACGCTAAGGCAAACGTAAGCCATTCCGCACTCAACGGCAATGTCT 2575
388 rGlyGlyAlaAspAla.....GlnGlyAspIleV 398
2576 CCCTAGCCGATAAGGCAGTATTCCATTTTGAAAACAGCCGCTTTACCGGA 2625
398 alAlaThr.....GluLeuProSerIleProGly 407
2626 AAAATCAGCGGCGGCAAGGATACGGCATTACACTTAAAAGACAGCGAATG 2675
408 ThrSerIleGlyProLeuAspValAlaLeu...AlaSerGlnAlaArgTr 423
2676 GACGCTGCCGTCGGGCACGGAATTAGGCAATTTAAACCTTGACAACGCCA 2725
423 pThr...GlyAlaThrArgAlaValAspSerLeuSerIleAspAsnAlaT 439
2726 CCATTACACTCAATTCCGCCTATCGACACGATGCGGCAGGCGCGCAAACC 2775
439 hrTrpValMetThrAsp..... 444
2776 GGCAGTGCGGCAGATGCGCCGCGCCGCCGTTCGCGCCGTTCCCTATTATC 2825
445 AsnSerAsnValGlyAlaLeuArgLeuAlaSerAspGlySer...ValAs 460
2826 CGTTACGCCGCCAACTTCGGCAGAATCCCGTTTCAACACGCTGACGGTAA 2875
460 pPheGlnGlnProAlaGluAla...GlyArgPheLysValLeuThrValA 476
2876 ACGGCAAATTGAACGGTCAGGGAACATTCCGCTTTATGTCGGAACTCTTC 2925
476 sn...ThrLeuAlaGlySerGlyLeuPheArgMetAsnValPheAlaAsp 491
2926 GGCTACCGCAGCGGCAAATTGAAGCTGGCGGAAAGTTCCGAAGGCACTTA 2975
492 LeuGlyLeuSerAspLysLeuValValMetGlnAspAlaSerGlyGlnHi 508
2976 CACCTTGGCTGTCAACAATACCGGCAACGAACCCGTAAGTCTCGAGCAAT 3025
508 sArgLeuTrpValArgAsnSerGlySerGluProAlaSerAlaAsnThrL 525
3026 TGACGGTAGTGGAAGGAAAAGACAACACACCGCTGTCCGAAAATCTTAAT 3075
525 euLeuLeuValGln.....ThrProLeuGlySerAlaAlaThr 537
3076 TTCACCCTGCAAAACGAA.....CACGTCGATGCCGGCGCATGGCGTTA 3119
538 PheThrLeuAlaAsnLysAspGlyLysValAspIleGlyThrTyrArgTy 554
3120 TCAGCTTATCCGCAAAGACGGCGAGTTCCGCCTGCATAATCCGGTCAAAG 3169
554 rArgLeu..... 556
3170 AACAAGAGCTTTCCGACAAACTCGGCAAGGCGGGAGAAACAGAGGCCGCC 3219
556 556
3220 TTGACGGCAAAACAGGCACAACTTGCCGCCAAACAACAGGCGGAAAAAGA 3269
557 .....AlaAlaAsnGl 560
3270 CAACGCGCAAAGCCTTGACGCGCTGATTGCGGCCGGGCGCAATGCCACCG 3319
560 yAsnGlyGlnTrp.....SerLeuValGlyAla..... 569
3320 AAAAGGCAGAAAGTGTTGCCGAACCGGCCCGGCAGGCAGGCGGGGAAAAT 3369
570 ..LysAlaProProAlaProLysProAlaProGlnProGlyProGln... 584
3370 GCCGGCATTATGCAGGCGGAGGAAGAGAAAAAACGGGTGCAGGCGGATAA 3419
584 584
3420 AGACACCGCCTTGGCGAAACAGCGCGAAGCGGAAACCCGGCCGGCTACCA 3469
585 .....ProProGlnPro..... 588
3470 CCGCCTTCCCCCGCGCCCGCCGCGCCCGCCGGGATTTGCCGCAACCGCAG 3519
589 .....ProGlnProGln 592
3520 CCCCAA.....CCGCAACCCCAACCGCAGCGCGACCTGATCAGCCG 3560
593 ProGluAlaProAlaProGlnProProAlaGlyArgGluLeu...SerAl 608
3561 TTATGCCAATAGCGGTTTGAGTGAATTTTCCGCCACGCTCAACAGCGTT. 3609
608 aAlaAlaAsnAlaAlaValAsnThrGlyGlyValGlyLeuAlaSerThrL 625
3610 .....TTCGCCGTACAGGACGAATTGGACCGCGTGTTTGCCGAAGACCGC 3654
625 euTrpTyrAlaGluSerAsnAlaLeuSerLysArgLeuGlyGluLeuArg 641
3655 CGCAAC.....GCCGTTTGGACAAGCGGCATCCGGGACACCAA 3692
642 LeuAsnProAspAlaGlyGlyAlaTrpGlyArgGlyPhe..... 654
3693 ACACTACCGTTCGCAAGATTTCCGCGCCTACCGCCAACAAACCGACCTGC 3742
655 .....AlaGlnArgGlnGlnLeuAsp.... 661
3743 GCCAAATCGGTATGCAGAAAAACCTCGGCAGCGGGCGCGTCGGCATCCTG 3792
661 661
3793 TTTTCGCACAACCGGACCGGAAACACCTTCGACGAC..... 3828
662 .....AsnArgAlaGlyArgArgPheAspGlnLysValAlaGlyPh 675
3829 .....GGCATCGGCAACTCGGCACGGCTTGCCCACGGTGCC.....G 3865
675 eGluLeuGlyAlaAspHisAlaValAlaValAlaGlyGlyArgTrpHisL 692
3866 TTTTCGGGCAATACGGCATCGGCAGGTTCGACATCGGCATCAGCGCGGGC 3915
692 euGlyGlyLeuAlaGlyTyrThrArgGlyAspArgGlyPheThrGlyAsp 708
3916 GCGGGTTTTAGTAGCGGCAGCCTT..... 3939
709 GlyGlyGlyHisThrAspSerValHisValGlyGlyTyrAlaThrTyrIl 725
3939 3939
725 eAlaAspSerGlyPheTyrLeuAspAlaThrLeuArgAlaSerArgLeuG 742
3940 .....TCAGACGGC.....ATCAGAGGCAAA 3960
742 luAsnAspPheLysValAlaGlySerAspGlyTyrAlaValLysGlyLys 758
3961 ATCCGCCGCCGCGTGCTGCATTACGGCATTCAGGCAAGATACCGCGCAGG 4010
759 TyrArgThr.....HisGlyValGlyAlaSerLeuGluAlaGl 771
4011 T.....TTCGGCGGATTCGGCATCGAACCGCACATCGGCG 4045
771 yArgArgPheThrHisAlaAspGlyTrpPheLeuGluProGlnAlaGluL 788
4046 CAACGCGCTATTTCGTCCAAAAAGCGGATTACCGATACGAAAAC...GTC 4092
788 euAlaValPheArgAlaGlyGlyGlyAlaTyrArgAlaAlaAsnGlyLeu 804
4093 AATATCGCCACCCCGGGCCTTGCATTCAACCGCTACCGCGCGGGCATTAA 4142

```
      ::::::          |||   :::           |||   |||:::::
  805 ArgValArgAspGluGlyGlySerSerValLeuGlyArgLeuGlyLeuGl 821

 4143 GGCAGATTATTCATTCAAACCGGCGCAACACATTTCCATCACGCCTTATT 4192
      ::::        ::::::   |||         ::::::   ||||||:
  821 uValGlyLysArgIleGluLeuAlaGlyGlyArgGlnValGlnProTyrI 838

 4193 TGAGCCTGTCCTATACCGATGCCGCTTCCGGCAAAGTCCGAACGCGCGTC 4242
      :::::   |||      :::      :::|||         :::::::::
  838 leLysAlaSerValLeuGlnGluPheAspGlyAlaGlyThrValHisThr 854

 4243 AATACCGCCGTATTGGCGCAGGATTTCGGCAAAACCCGCAGTGCGGAATG 4292
      |||      :::         ::::::      ||||||   ||||||
  855 AsnGlyIleAlaHisArgThrGluLeuArgGlyThrArg...AlaGluLe 870

 4293 GGGCGTAAACGCCGAAATCAAAGGTTTCACGCTGTCCCTCCACGCTGCCG 4342
       |||                            |||::::::   ||||||
  870 uGly.....................LeuGlyMetAlaAlaAlaL 878

 4343 CCGCCAAGGGGCCGCAATTGGAAGCG......CAGCACAGCGCGGGCATC 4386
      ::::::|||   :::|||   |||      ::::::|||   |||
  878 euGlyArgGlyHisSerLeuTyrAlaSerTyrGluTyrSerLysGlyPro 894

 4387 AAATTAGGCTACCGCTGG 4404
      ||||||:::      |||
  895 LysLeuAlaMetProTrp 900

seq_name: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:US-08-038-682-2

seq_documentation_block:
; Sequence 2, Application US/08038682
; Patent No. 5549897
;  GENERAL INFORMATION:
;     APPLICANT:  BARENKAMP, STEPHEN J
;     APPLICANT:  ST. GEME III, JOSEPH W
;     TITLE OF INVENTION:  HIGH MOLECULAR WEIGHT SURFACE PROTEINS
;     TITLE OF INVENTION:  OF NON-TYPEABLE HAEMOPHILUS
;     NUMBER OF SEQUENCES:  8
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Shoemaker and Mattare, Ltd
;       STREET:  2001 Jefferson Davis Hwy., 1203 Crystal Plaza
;       STREET:  Bldg. 1
;       CITY:  Arlington
;       STATE:  Virginia
;       COUNTRY:  U.S.A.
;       ZIP:  22202-0286
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/038,682
;       FILING DATE:  16-MAR-1993
;       CLASSIFICATION:  424
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  BERKSTRESSER, JERRY W
;       REGISTRATION NUMBER:  22,651
;       REFERENCE/DOCKET NUMBER:  1038-293
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (703) 415-0810
;       TELEFAX:  (703) 415-0813
;   INFORMATION FOR SEQ ID NO:  2:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1536 amino acids
;       TYPE:  amino acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  protein
US-08-038-682-2


alignment_scores:

                Quality:  265.00          Length:   1297
                  Ratio:  0.438             Gaps:     73
  Percent Similarity:  46.646   Percent Identity: 21.126

alignment_block:
 US-09-303-518D-653 x US-08-038-682-2   ..

 Align seg 1/1  to: US-08-038-682-2  from: 1  to: 1536

    178 AATAAAGGCAAGTTTGCAGTCGGGGCGAAAGATATTGAGGTTTACAACAA 227
        |||  |||   ::::::||||||   ||||||   :::|||
    199 AsnHisGlyLeuIleThrValGly...LysAspGlyServal........ 211

    228 AAAAGGGGAGTTGGTCGGCAAATCGATGACGAAAGCCCCGATGATTGATT 277
              :::|||:::|||   ::::::   :::
    212 .......AsnLeuIleGlyGlyLysValLysAsnGlu............. 221

    278 TTTCTGTGGTATCGCGTAACGGC......GTGGCGGCATTGGCGGGCGAT 321
          :::|||:::|||   ||||||      ::::::   ||||||||||:::
    222 ..GlyValIleSerValAsnGlyGlySerIleSerLeuLeuAlaGlyGln 237

    322 CAA.....................................TATATTGTGAG 335
        :::                                     |||   :::::
    238 LysIleThrIleSerAspIleIleAsnProThrIleThrTyrSerIleAl 254

    336 CGTGGCACATAACGGCGGCTATAACAATGTTGATTTTGGTGCCGAGGGAA 385
        ::::   :::|||   :::   |||      |||:::   |||:::|||:
    254 aAlaProGluAsnGluAlaValAsnLeuGlyAspIlePheAlaLysGlyG 271

    386 GCAAT.......................CCCGATCAGCACCGCTTTTCT 411
        ::|||                           :::|||   :::::|||
    271 lyAsnIleAsnValArgAlaAlaThrIleArgAsnGlnGlyLysLeuSer 287

    412 TACCAAATTGTGAAAAGAAATAATTATAAAGCAGGGACTAACGGCCATCC 461
        :::   |||:::::::::      |||:::|||
    288 AlaAspSerValSerLysAsp......LysSerGly.............. 297

    462 TTATGGCGGCGATTATCATATGCCGCGTTTGCACAAATTTGTCACAGATG 511
                                        ::::::|||   :::|
    298 ................................AsnIleValLeuSerA 303

    512 CAGAACCTGTTGAG.................ATGACCAGTTATATGGAT 543
        ||:::      |||                 :::::::::       :::
    303 laLysGluGlyGluAlaGluIleGlyGlyValIleSerAlaGlnAsnGln 319

    544 GGGTGGAAATACGCTGATTTAAATAAATACCCTGATCGTGTT......CG 587
            |||   :::   |||             |||:::|||      ::
    320 GlnAlaLysGlyGlyLysLeuMetIleThrGlyAspLysValThrLeuLy 336

    588 AATCGGAGCA.........................GGCAGACAATATT 610
        :   ||||||                         |||:::   |||
    336 sThrGlyAlaValIleAspLeuSerGlyLysGluGlyGlyGluThrTyrL 353

    611 GGCGGTCTGATGAA...GACGAACCCAATAAC.................. 639
          :::||||||      |||   :::|||
    353 euGlyGlyAspGluArgGlyGluGlyLysAsnGlyIleGlnLeuAlaLys 369

    640 ..............CGCGAAAGTTCATATCATATTGCAAGC........ 666
                     :::   |||:::   ::::::::::::
    370 LysThrSerLeuGluLysGlySerThrIleAsnValSerGlyLysGluLy 386

    667 .........GCATATTCTTGG............CTCGTCGGTGGCAATA 694
                |||       |||            |||:::   ||||||
    386 sGlyGlyArgAlaIleValTrpGlyAspIleAlaLeuIleAspGlyAsnI 403

    695 CCTTTGCACAAAATGGATCAGGTGGTGGCACAGTCAACTTAGGTAGCGAA 744
          ||||||   |||||||||
    403 leAsnAlaGln...GlySerGly........................... 409

    745 AAAATTAAACATAGCCCATATGGTTTTTTACCAACAGGAGGCTCATTTGG 794
                                  :::      |||||||||
    410 ....................AspIleAlaLysThrGlyGlyPheValGl 419
```

```
 959 .................ThrIleIleSerGlyAsnIleThrAsnLysAsn 969

2107 GGCGATTGGCATTTAAGCAATCACGCC............CAAGCAGTTTT 2144
     ||||||   ::::::::::|||:::::::              :::   :::
 970 GlyAspLeuAsnIleThrAsnGluGlySerAspThrGluMetGlnIleGl 986

2145 CGGTGTCGCACCGCATCAAAGCCACACAATCTGTACACGTTCGGACTGGA 2194
      |||   :::   :::::::::   ::::::          ||||||
 986 yGlyAspValSerGlnLysGluGlyAsnLeuThrIleSerSerAspLysI 1003

2195 CGGGTCTGACAAGTTGTACCGAAAAAACCATTACC...GACGATAAAGTG 2241
       :::::::|||            |||   ||||||            |||
1003 leAsnIleThr............LysGlnIleThrIleLysAlaGlyVal 1015

2242 ATTGCTTCATTGAGCAAGACCGACATCAGAGGCAATGTCAGCCTTGCCGA 2291
       ::::::   |||   :::|||      :::|||::::::|||:::
1016 AspGlyGluAsnSerAspSerAspAlaThrAsnAsnAlaAsnLeuThrIl 1032

2292 TCACGCT.....................CATTTAAATCTCACAGGACTT. 2319
       :::                      |||||:::::|||:::
1032 eLysThrLysGluLeuLysLeuThrGlnAspLeuAsnIleSerGlyPheA 1049

2320 ................................GCCACACTCAACGGC 2334
                                     |||::::::::::
1049 snLysAlaGluIleThrAlaLysAspGlySerAspLeuThrIleGlyAsn 1065

2335 AATCTTAGTGCAGGCGGAGACACGCACTATACGGTTACGCGCAACGCCAC 2384
      :::   ||||||   |||   :::          ||||||   |||   ::
1066 ThrAsnSerAlaAspGlyThrAsnAlaLysLysValThrPheAsnGlnVa 1082

2385 CCAAAACGGCAACCTCAGCCTCGTGGGCAATGCCCAAGCAACATTTAATC 2434
     ::::::::::::::::|||      |||:::      :::      ::::
1082 lLysAspSerLysIleSerAlaAspGlyHisLysValThrLeuHisSerL 1099

2435 AAGCCACATTAAACGGCAACACATCGGCTTCGGACAATGCTTCATTTAAT 2484
     :::::      :::|||:::::::::   :::::::::::::|||   |||
1099 ysValGluThrSerGlySerAsnAsnAsnThrGluAspSerSerAspAsn 1115

2485 CTAAGCAACAACGCCGTACAAAACGGCAGTCTGACGCTTTCCGACAACGC 2534
       ::::::   :::::::::      ::::::|||:::::::::|||
1116 AsnAlaGlyLeuThrIleAspAlaLysAsnValThrValAsnAsnAsnIl 1132

2535 TAAGGCAAAC......GTAAGCCATTCCGCACTCAACGGCAATGTC.... 2574
       ::::::      ||||||   ||||||   :::|||::::::
1132 eThrSerHisLysAlaValSerIleSerAlaThrSerGlyGluIleThrT 1149

2575 ..........TCCCTAGCCGATAAGGCAGTATTCCATTTTGAAAACAGC 2613
                :::      :::   ::::::   :::   |||   :::
1149 hrLysThrGlyThrThrIleAsnAlaThrThrGlyAsnValGluIleThr 1165

2614 CGCTTTACCGGAAAAATCAGCGGCGGCAAGGATACG.........GCATT 2654
           ||||||:::|||   ||||||   ::::::          :::::
1166 AlaGlnThrGlySerIleLeuGlyGlyIleGluSerSerSerGlySerVa 1182

2655 ACACTTAAAAGACAGCGAATGGACGCTGCCGTCGGGCACGGAATTAGGCA 2704
     :   |||       :::|||   :::|||                 ::::::|
1182 lThrLeuThrAlaThrGluGlyAlaLeu............AlaValSerA 1195

2705 ATTTAAACCTTGACAACGCCACCATTACACTCAATTCCGCCTATCGACAC 2754
     ||:::::   :::::::::|||:::|||   ||||||
1195 snIleSerGlyAsnThrValThrValThrAlaAsnSer............ 1207

2755 GATGCGGCAGGCGCGCAAACCGGCAGTGCGGCAGATGCGCCGCGCCGCCG 2804
           ||||||   |||      |||::::::
1208 .........GlyAlaLeuThrThrLeuAlaGlySer.............. 1216

2805 TTCGCGCCGTTCCCTATTATCCGTTACGCCGCCAACTTCGGCAGAATCCC 2854
      :::   ::::::      |||||||||      ||||||::::::|||
1217 .ThrIleLysGlyThrGluSerValThr......ThrSerSerGlnSerG 1231

2855 GTTTCAACACGCTGACGGTAAACGGCAAATTGAACGGTCAGGGAACA... 2901
                     ::::::|||   ::::::|||   ||||||
1231 ly............AspIleGlyGlyThrIleSerGly...GlyThrVal 1242
```

```
2902 ...TTCCGCTTTATGTCGGAACTCTTCGGCTACCGCAGCGGCAAATTGAA 2948
           :::      ::::::|||              ::::::|||:::||
1243 GluValLysAlaThrGluSerLeuThrThrGlnSerAsnSerLysIleLy 1259

2949 GCTGGCGGAAAGTTCCGAAGGCACTTACACCTTGGCTGTC.......... 2988
     |   |||   ::::::|||::::::   |||   |||:::
1259 s...AlaThrThrGlyGluAlaAsnValThrSerAlaThrGlyThrIleG 1275

2989 ..............AACAATACCGGCAACGAACCCGTAAGTCTCGAGCAA 3024
                     :::|||||   |||      ::::::      :::
1275 lyGlyThrIleSerGlyAsnThrValAsnValThrAlaAsnAlaGlyAsp 1291

3025 TTGACGGTA.......................GTGGAAGGAAAAGACAA 3050
     |||||||||                      :::||||||      ::
1292 LeuThrValGlyAsnGlyAlaGluIleAsnAlaThrGluGlyAlaAlaTh 1308

3051 CACACCGCTGTCCGAAAATCTTAATTTCACCCTGCAAAACGAACACGTCG 3100
     :         |||::::::          |||      ::::::|||:::
1308 rLeuThrThrSerSerGlyLysLeuThrThrGluAlaSerSerHisIleT 1325

3101 ATGCCGGCGCATGGCGTTATCAGCTTATCCGCAAAGACGGCGAGTTCCGC 3150
       ::: |||   |||   ::::::|||||||||:::   :::|||   ||
1325 hrSer.AlaLysGlyGlnValAsnLeuSerAlaGlnAspGlySerValAl 1341

3151 CTGCATAATCCGGTCAAAGAACAAGAGCTTTCCGACA 3187
     |      |||:::::::   |||   ::::::   |||
1341 aGlySerIleAsnAlaAlaAsnValThrLeuAsnThr 1353
```

seq_name: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:US-08-302-832-2

```
seq_documentation_block:
; Sequence 2, Application US/08302832
; Patent No. 5603938
;  GENERAL INFORMATION:
;     APPLICANT:  Barenkamp, Stephen J
;     TITLE OF INVENTION:  High Molecular Weight Surface Proteins
;     TITLE OF INVENTION:  of No. 5603938-Typeable Haemophilus
;     NUMBER OF SEQUENCES:  8
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Shoemaker and Mattare, Ltd.
;       STREET:  2001 Jefferson Davis Hwy., 1203 Crystal Plaza
;       STREET:  Bldg. 1
;       CITY:  Arlington
;       STATE:  Virginia
;       COUNTRY:  U.S.A.
;       ZIP:  22202-0286
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/302,832
;       FILING DATE:  16-SEP-1994
;       CLASSIFICATION:  435
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  GB 9205704.1
;       FILING DATE:  16-MAR-1992
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US pct/us93/02166
;       FILING DATE:  16-MAR-1993
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Berkstresser, Jerry W
;       REGISTRATION NUMBER:  22,651
;       REFERENCE/DOCKET NUMBER:  1038-404
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (703) 415-0810
;       TELEFAX:  (703) 415-0813
;  INFORMATION FOR SEQ ID NO:  2:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1536 amino acids
;       TYPE:  amino acid
```

```
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-302-832-2


alignment_scores:
            Quality:  265.00              Length:   1297
              Ratio:  0.438                Gaps:     73
 Percent Similarity:  46.646   Percent Identity: 21.126

alignment_block:
 US-09-303-518D-653 x US-08-302-832-2   ..

 Align seg 1/1  to: US-08-302-832-2  from: 1  to: 1536

    178 AATAAAGGCAAGTTTGCAGTCGGGGCGAAAGATATTGAGGTTTACAACAA 227
        |||   |||   :::::::||||||   ||||||   :::|||
    199 AsnHisGlyLeuIleThrValGly...LysAspGlySerVal........ 211

    228 AAAAGGGGAGTTGGTCGGCAAATCGATGACGAAAGCCCCGATGATTGATT 277
              :::|||:::|||   ::::::   :::
    212 .......AsnLeuIleGlyGlyLysValLysAsnGlu............. 221

    278 TTTCTGTGGTATCGCGTAACGGC......GTGGCGGCATTGGCGGGCGAT 321
          :::|||:::|||   ||||||      ::::::   ||||||||||:::
    222 ..GlyValIleSerValAsnGlyGlySerIleSerLeuLeuAlaGlyGln 237

    322 CAA.....................................TATATTGTGAG 335
        :::                                    |||   :::::
    238 LysIleThrIleSerAspIleIleAsnProThrIleThrTyrSerIleAl 254

    336 CGTGGCACATAACGGCGGCTATAACAATGTTGATTTTGGTGCGGAGGGAA 385
        ::::   :::|||   :::   |||      |||:::   |||:::|||:
    254 aAlaProGluAsnGluAlaValAsnLeuGlyAspIlePheAlaLysGlyG 271

    386 GCAAT.......................CCCGATCAGCACCGCTTTTCT 411
        ::|||                                :::|||   ::::::|||
    271 lyAsnIleAsnValArgAlaAlaThrIleArgAsnGlnGlyLysLeuSer 287

    412 TACCAAATTGTGAAAAGAAATAATTATAAAGCAGGGACTAACGGCCATCC 461
          :::   |||:::::::::      |||:::|||
    288 AlaAspSerValSerLysAsp......LysSerGly.............. 297

    462 TTATGGCGGCGATTATCATATGCCGCGTTTGCACAAATTTGTCACAGATG 511
                                           :::::::|||   ::::|
    298 ..................................AsnIleValLeuSerA 303

    512 CAGAACCTGTTGAG.................ATGACCAGTTATATGGAT 543
        ||:::      |||                 :::::::::      :::
    303 laLysGluGlyGluAlaGluIleGlyGlyValIleSerAlaGlnAsnGln 319

    544 GGGTGGAAATACGCTGATTTAAATAAATACCCTGATCGTGTT......CG 587
              |||   :::   |||             |||:::|||      ::
    320 GlnAlaLysGlyGlyLysLeuMetIleThrGlyAspLysValThrLeuLy 336

    588 AATCGGAGCA..........................GGCAGACAATATT 610
        :   ||||||                         |||:::   |||
    336 sThrGlyAlaValIleAspLeuSerGlyLysGluGlyGlyGluThrTyrL 353

    611 GGCGGTCTGATGAA...GACGAACCCAATAAC.................. 639
          :::|||||||      |||   :::|||
    353 euGlyGlyAspGluArgGlyGluGlyLysAsnGlyIleGlnLeuAlaLys 369

    640 ..............CGCGAAAGTTCATATCATATTGCAAGC........ 666
                     :::   |||:::   ::::::::::::
    370 LysThrSerLeuGluLysGlySerThrIleAsnValSerGlyLysGluLy 386

    667 ..........GCATATTCTTGG............CTCGTCGGTGGCAATA 694
                  |||      |||            |||:::   ||||||
    386 sGlyGlyArgAlaIleValTrpGlyAspIleAlaLeuIleAspGlyAsnI 403

    695 CCTTTGCACAAAATGGATCAGGTGGTGGCACAGTCAACTTAGGTAGCGAA 744

        ||||||   |||||||||
403 leAsnAlaGln...GlySerGly.......................... 409

745 AAAATTAAACATAGCCCATATGGTTTTTTACCAACAGGAGGCTCATTTGG 794
                              :::      |||||||||
410 ....................AspIleAlaLysThrGlyGlyPheValGl 419

795 CGACAGTGGCTCACCAATGTTTATCTATGATGCCCAAAAGCAAAAGTGGT 844
      ||||||      :::||||||   |||
419 uThrSerGlyHisAspLeuPheIleLysAsp.................... 429

845 TAATTAATGGGGTATTGCAAACAGGCAACCCCTATATAGGAAAAAGCAAT 894
      |||:::::::::::::::::
430 .....AsnAlaIleValAspAla........................... 435

895 GGCTTCCAGCTAGTTCGTAAAGATTGGTTCTATGATGAAATCTTTGCTGG 944
                        |||:::|||:::   |||
436 ..................LysGluTrpLeuLeuAsp............Ph 442

945 AGATACCCATTCAGTATTCTACGAACCACATCAAAATGGGAAATACTTTT 994
     |||:::   |||:::      |||         :::::::::
442 eAspAsnValSerIleAsnAlaGluThrAlaGlyArgSerAsnThrSerG 459

995 TTAACGACAATAATAATGGCGCAGGAAAAATCGATGCCAAACATAAACAC 1044
      :::|||:::   :::|||:::|||                    :::
459 luAspAspGluTyrThrGlySerGly..................AsnSer 469

1045 TATTCTCTACCTTATAGATTAAAAACACGAACCGTTCAATTGTTTAATGT 1094
       |||   |||   |||   |||   :::|||:::   |||   |||::
470 AlaSerThrProLysArgAsnLysGluLysThrThr...LeuThrAsnTh 485

1095 TTCTTTATCCGAG.............................ACAGCAA 1114
     ::::|||::::::                            |||||||:
485 rThrLeuGluSerIleLeuLysLysGlyThrPheValAsnIleThrAlaA 502

1115 GAGAACCTGTTTAT..................................... 1128
     :::::   :::|||
502 snGlnArgIleTyrValAsnSerSerIleAsnLeuSerAsnGlySerLeu 518

1129 ..................CATGCTGCAGGTGGGGTCAACAGTTATCGACC 1160
                        :::::::::||||||||||
519 ThrLeuTrpSerGluGlyArgSerGlyGlyGlyVal.............. 530

1161 CAGACTGAATAAT.........................GGAGAAAATA 1183
      ::::::|||||                          |||   |||:
531 .GluIleAsnAsnAspIleThrThrGlyAspAspThrArgGlyAlaAsnL 547

1184 TTTCCTTTATTGACAAAGGAAAAGGTGAATTGATACTTACCAGCAACATC 1233
     ::::::::   :::   |||      :::      :::   :::|||||||
547 euThrIleTyrSerGlyGlyTrpValAsp......ValHisLysAsnIle 561

1234 AACCAAGGCGCGGGCGGTTTGTATTTTGAGGGTAATTTTACGGTCTCGCC 1283
     :::   ||||||               :::|||||||:::::::::::::
562 SerLeuGlyAla..............GlnGlyAsnIleAsnIleThrAl 573

1284 TAAA......................AACAACGAAACGTGGCAAGGCG 1309
      |||                     :::|||::::::      |||
573 aLysGlnAspIleAlaPheGluLysGlySerAsnGlnValIleThrGlyG 590

1310 CGGGCGTTCATATCAGTGATGGCAGTACCGTTACTTGGAAAGTAAACGGC 1359
       |||:::   |||:::::::|||:::          ::::::   |||:::
590 lnGlyThr...IleThrSerGlyAsnGlnLysGlyPheArgPheAsnAsn 605

1360 GTGGCAAACGACCGCCTGTCCAAAATCGGCAAAGGCACGCTGCTGGTTCA 1409
     |||:::         |||:::      |||::::||   |||
606 ValSer.........LeuAsnGlyThrGlySerGly...LeuGlnPheTh 618

1410 AGCCAAAGGGGAAAAC.....................CAAGGCTCGG 1435
       :::|||      |||                    :::|||::::
618 rThrLysArgThrAsnLysTyrAlaIleThrAsnLysPheGluGlyThrL 635

1436 TCAGCGTGGGCGACGGTAAAGTCATCTTA.................GAT 1467
     ::::::::   :::||||||||||   :::                 :::
```

```
 635 euAsnIle...SerGlyLysValAsnIleSerMetValLeuProLysAsn 650

1468 CAGCAGGCGGACGATCAAGGCAAAAAACAAGCCTTT..............  1503
     :::::::::   |||:::   |||   :::::::::
 651 GluSerGlyTyrAspLysPheLysGlyArgThrTyrTrpAsnLeuThrSe  667

1504 ..........AGTGAAATCGGC............TTGGTCAGCGGCAGGG 1531
             ||||||   |||          :::::::::||||
 667 rLeuAsnValSerGluSerGlyGluPheAsnLeuThrIleAspSerArgG  684

1532 GGACG.........................................GTG 1539
     ||:::                                        :::
 684 lySerAspSerAlaGlyThrLeuThrGlnProTyrAsnLeuAsnGlyIle  700

1540 CAACTGAATGCCGATAATCAGTTCAACCCCGACAAA.............. 1575
     :::::::|||   |||:::   ||||||   ::::::
 701 SerPheAsnLysAspThrThrPheAsnValGluArgAsnAlaArgValAs  717

1576 ..........................................CTCT 1579
                                               |
 717 nPheAspIleLysAlaProIleGlyIleAsnLysTyrSerSerLeuAsnT  734

1580 ATTTCGGCTTTCGCGGCGGACGTTTGGATTTGAACGGGCATTCGCTTTCG 1629
     ||   :::|||:::|||:::              :::|||   |||::::::
 734 yrAlaSerPheAsnGlyAsnIleSerValSerGlyGlyGlySerValAsp  750

1630 TTCCACCGCATT.........CAAAATACCGATGAAGGGGCGATGATTGT 1670
     |||      :::      :::|||:::::      ::::::::::
 751 PheThrLeuLeuAlaSerSerSerAsnValGlnThrProGlyValValIl  767

1671 CAACCACAAT............................CAAGACAAAG 1690
     :|||   :::                           :::   |||
 767 eAsnSerLysTyrPheAsnValSerThrGlySerSerLeuArgPheLysT  784

1691 AATCCACCGTTACCATTACAGGC.........AATAAAGATATTACT... 1728
       |||      |||   ||||||          :::||||||:::|||
 784 hrSerGlySerThrLysThrGlyPheSerIleGluLysAspLeuThrLeu  800

1729 ...ACAACCGGCAATAAC.............................. 1743
        :::||||||:::|||
 801 AsnAlaThrGlyGlyAsnIleThrLeuLeuGlnValGluGlyThrAspGl  817

1744 ..........AACAACTTGGATAGCAAAAAAGAAATTGCCTACAACGGT. 1782
             :::::::::   :::||||||::|||::::::::::|||
 817 yMetIleGlyLysGlyIleValAlaLysLysAsnIleThrPheGluGlyG  834

1783 ........TGGTTTGGCGAGAAAGATGCA...ACCAAAACGAACGGGCGG 1821
             ||||||:::::   |||   |||:::   :::|||:::
 834 lyAsnIleThrPheGlySerArgLysAlaValThrGluIleGluGlyAsn  850

1822 CTCAATCTGAATTACCAACCGGAAGAAGCGGATCGCACTTTA........ 1863
     :::::::::|||          ::::::|||:::   ||||||
 851 ValThrIleAsn.........AsnAsnAlaAsnValThrLeuIleGlySe  864

1863 ..................................................  1863

 864 rAspPheAspAsnHisGlnLysProLeuThrIleLysLysAspValIleI  881

1864 ..............CTGCTTTCCGGCGGA......ACAAATTTAAACGGC 1893
                 |||   :::||||||      :::|||:::   |||
 881 leAsnSerGlyAsnLeuThrAlaGlyGlyAsnIleValAsnIleAlaGly  897

1894 AATATCACG...CAAACAAACGGCAAACTGTTTTTCAGCGGCA....... 1933
     |||:::|||   ::::::||| ||||||   |||:::|||
 898 AsnLeuThrValGluSerAsn.AlaAsn...PheLysAlaIleThrAsnP  913

1934 ........................GACCGACACCGCACGCCTACAATC 1957
                             |||:::      :::   :::|||
 913 heThrPheAsnValGlyGlyLeuPheAspAsnLysGlyAsnSerAsnIle  929

1958 ATTTAGGAAGCGGGTGGTCAAAAATGGAAGGTATCCCACAAGGAGAAATC 2007
             ||||||   :::::::|||         ::::::|||::
 930 SerIleAlaLysGlyGlyAlaArgPheLysAspIleAspAsnSerLysAs  946

2008 GTGTGGGACAACGATTGGATCGACCGC.ACATTTAAAGCGGAAAACTTCC 2056
     :   :::   |||   :::|||:::   |||::::::
  946 nLeuSerIleThrThrAsnSerSerSerThrTyrArg.............  958

2057 ATATTCAGGGCGGACAAGCGGTGGTTTCCCGCAATGTTGCCAAAGTGGAA 2106
                    :::::::::|||   |||::::::::::   :::
  959 .................ThrIleIleSerGlyAsnIleThrAsnLysAsn  969

2107 GGCGATTGGCATTTAAGCAATCACGCC............CAAGCAGTTTT 2144
     ||||||   :::::::::|||::::::          :::   :::
  970 GlyAspLeuAsnIleThrAsnGluGlySerAspThrGluMetGlnIleGl  986

2145 CGGTGTCGCACCGCATCAAAGCCACACAATCTGTACACGTTCGGACTGGA 2194
      |||   :::   :::::::::   ::::::         ||||||
  986 yGlyAspValSerGlnLysGluGlyAsnLeuThrIleSerSerAspLysI 1003

2195 CGGGTCTGACAAGTTGTACCGAAAAAACCATTACC...GACGATAAAGTG 2241
       ::::::|||            |||   ||||||            |||
 1003 leAsnIleThr............LysGlnIleThrIleLysAlaGlyVal 1015

2242 ATTGCTTCATTGAGCAAGACCGACATCAGAGGCAATGTCAGCCTTGCCGA 2291
       ::::::   |||   :::|||      :::|||:::::::|||:::
 1016 AspGlyGluAsnSerAspSerAspAlaThrAsnAsnAlaAsnLeuThrIl 1032

2292 TCACGCT.....................CATTTAAATCTCACAGGACTT. 2319
        :::                      ||||||:::::|||:::
 1032 eLysThrLysGluLeuLysLeuThrGlnAspLeuAsnIleSerGlyPheA 1049

2320 ................................GCCACACTCAACGGC 2334
                                     |||::::::::::
 1049 snLysAlaGluIleThrAlaLysAspGlySerAspLeuThrIleGlyAsn 1065

2335 AATCTTAGTGCAGGCGGAGACACGCACTATACGGTTACGCGCAACGCCAC 2384
     :::   ||||||   |||   :::          ||||||   |||   ::
 1066 ThrAsnSerAlaAspGlyThrAsnAlaLysLysValThrPheAsnGlnVa 1082

2385 CCAAAACGGCAACCTCAGCCTCGTGGGCAATGCCCAAGCAACATTTAATC 2434
     :::::::::::::::::|||      |||:::      :::      ::::
 1082 lLysAspSerLysIleSerAlaAspGlyHisLysValThrLeuHisSerL 1099

2435 AAGCCACATTAAACGGCAACACATCGGCTTCGGACAATGCTTCATTTAAT 2484
     :::::      :::|||::::::::::   :::::::::::::|||   |||
 1099 ysValGluThrSerGlySerAsnAsnAsnThrGluAspSerSerAspAsn 1115

2485 CTAAGCAACAACGCCGTACAAAACGGCAGTCTGACGCTTTCCGACAACGC 2534
     ::::::   ::::::::::      ::::::|||::::::::::|||
 1116 AsnAlaGlyLeuThrIleAspAlaLysAsnValThrValAsnAsnAsnIl 1132

2535 TAAGGCAAAC......GTAAGCCATTCCGCACTCAACGGCAATGTC.... 2574
        ::::::      ||||||   ||||||   :::|||::::::
 1132 eThrSerHisLysAlaValSerIleSerAlaThrSerGlyGluIleThrT 1149

2575 ..........TCCCTAGCCGATAAGGCAGTATTCCATTTTGAAAACAGC 2613
                :::      :::   ::::::   :::   |||   :::
 1149 hrLysThrGlyThrThrIleAsnAlaThrThrGlyAsnValGluIleThr 1165

2614 CGCTTTACCGGAAAAATCAGCGGCGGCAAGGATACG.........GCATT 2654
          ||||||:::|||   ||||||   ::::::         :::::
 1166 AlaGlnThrGlySerIleLeuGlyGlyIleGluSerSerSerGlySerVa 1182

2655 ACACTTAAAAGACAGCGAATGGACGCTGCCGTCGGGCACGGAATTAGGCA 2704
     :   |||      :::|||   :::|||                  ::::::|
 1182 lThrLeuThrAlaThrGluGlyAlaLeu............AlaValSerA 1195

2705 ATTTAAACCTTGACAACGCCACCATTACACTCAATTCCGCCTATCGACAC 2754
     ||:::::::   ::::::::::|||:::|||   ||||||
 1195 snIleSerGlyAsnThrValThrValThrAlaAsnSer............ 1207

2755 GATGCGGCAGGCGCGCAAACCGGCAGTGCGGCAGATGCGCCGCGCCGCCG 2804
             ||||||   |||      |||::::::
 1208 .........GlyAlaLeuThrThrLeuAlaGlySer.............. 1216
```

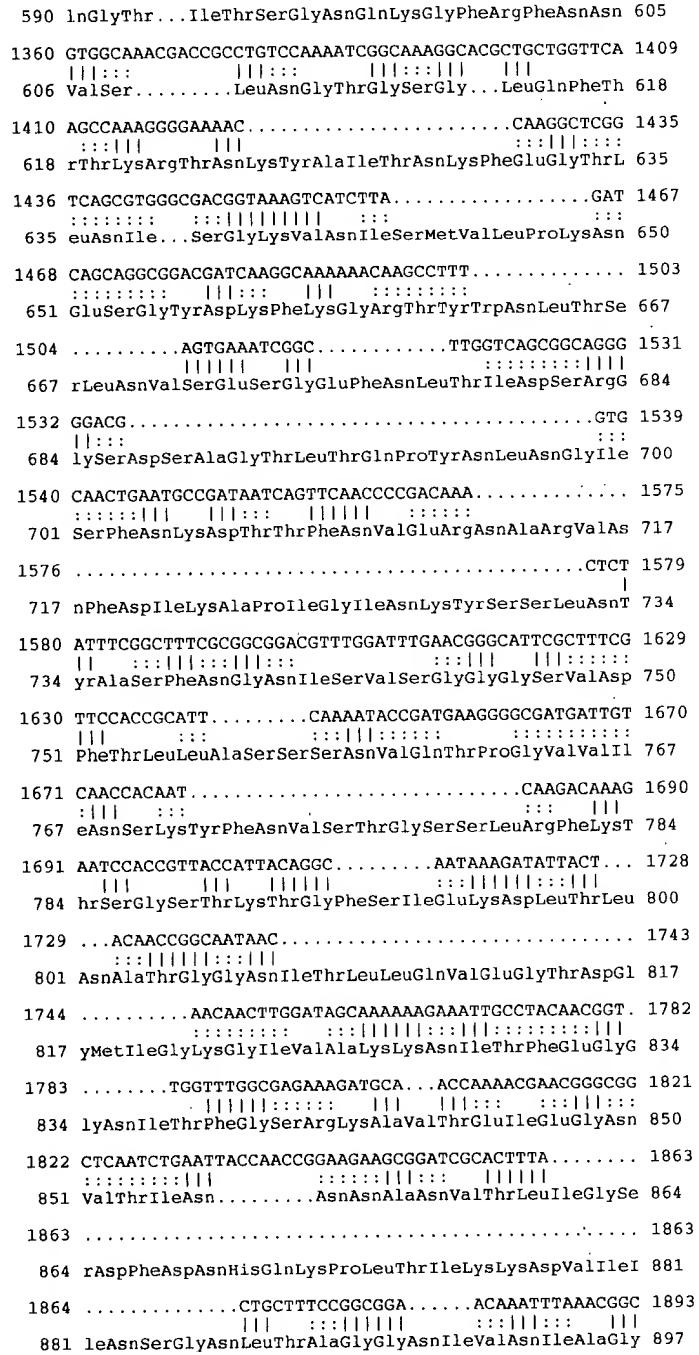
```
 2805 TTCGCGCCGTTCCCTATTATCCGTTACGCCGCCAACTTCGGCAGAATCCC 2854
      :::   ::::::      |||||||||      ||||||::::::|||
 1217 .ThrIleLysGlyThrGluSerValThr......ThrSerSerGlnSerG 1231

 2855 GTTTCAACACGCTGACGGTAAACGGCAAATTGAACGGTCAGGGAACA... 2901
                      ::::::|||   :::::::|||   ||||||
 1231 ly............AspIleGlyGlyThrIleSerGly...GlyThrVal 1242

 2902 ...TTCCGCTTTATGTCGGAACTCTTCGGCTACCGCAGCGGCAAATTGAA 2948
         :::      :::::::|||              ::::::|||:::||
 1243 GluValLysAlaThrGluSerLeuThrThrGlnSerAsnSerLysIleLy 1259

 2949 GCTGGCGGAAAGTTCCGAAGGCACTTACACCTTGGCTGTC.......... 2988
      |   |||   :::::::|||::::::   |||   |||:::
 1259 s...AlaThrThrGlyGluAlaAsnValThrSerAlaThrGlyThrIleG 1275

 2989 ..............AACAATACCGGCAACGAACCCGTAAGTCTCGAGCAA 3024
                     :::||||||   |||      ::::::      :::
 1275 lyGlyThrIleSerGlyAsnThrValAsnValThrAlaAsnAlaGlyAsp 1291

 3025 TTGACGGTA.......................GTGGAAGGAAAAGACAA 3050
      |||||||||                      :::||||||      ::
 1292 LeuThrValGlyAsnGlyAlaGluIleAsnAlaThrGluGlyAlaAlaTh 1308

 3051 CACACCGCTGTCCGAAAATCTTAATTTCACCCTGCAAAACGAACACGTCG 3100
      :          |||::::::         |||      ::::::|||:::
 1308 rLeuThrThrSerSerGlyLysLeuThrThrGluAlaSerSerHisIleT 1325

 3101 ATGCCGGCGCATGGCGTTATCAGCTTATCCGCAAAGACGGCGAGTTCCGC 3150
        ::: |||   |||   ::::::||||||||||:::   :::|||   ||
 1325 hrSer.AlaLysGlyGlnValAsnLeuSerAlaGlnAspGlySerValAl 1341

 3151 CTGCATAATCCGGTCAAAGAACAAGAGCTTTCCGACA 3187
      |      |||::::::   |||   ::::::   |||
 1341 aGlySerIleAsnAlaAlaAsnValThrLeuAsnThr 1353

seq_name: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:US-08-530-198-2

seq_documentation_block:
; Sequence 2, Application US/08530198
; Patent No. 5869065
;  GENERAL INFORMATION:
;    APPLICANT:  BARENKAMP, STEPHEN J
;    APPLICANT:  ST. GEME III, JOSEPH W
;    TITLE OF INVENTION:  HIGH MOLECULAR WEIGHT SURFACE PROTEINS
;    TITLE OF INVENTION:  OF NON-TYPEABLE HAEMOPHILUS
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Shoemaker and Mattare, Ltd
;      STREET:  2001 Jefferson Davis Hwy., 1203 Crystal Plaza
;      STREET:  Bldg. 1
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22202-0286
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/530,198
;      FILING DATE:  13-DEC-1995
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BERKSTRESSER, JERRY W
;      REGISTRATION NUMBER:  22,651
;      REFERENCE/DOCKET NUMBER:  JWB-1186
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 415-0810
;      TELEFAX:  (703) 415-0813
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1536 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-530-198-2


alignment_scores:
            Quality:  265.00              Length:   1297
              Ratio:  0.438                 Gaps:     73
 Percent Similarity:  46.646   Percent Identity: 21.126

alignment_block:
 US-09-303-518D-653 x US-08-530-198-2   ..

 Align seg 1/1  to: US-08-530-198-2  from: 1  to: 1536

     178 AATAAAGGCAAGTTTGCAGTCGGGGCGAAAGATATTGAGGTTTACAACAA 227
         |||   |||   ::::::|||||||   ||||||   :::|||
     199 AsnHisGlyLeuIleThrValGly...LysAspGlySerVal........ 211

     228 AAAAGGGGAGTTGGTCGGCAAATCGATGACGAAAGCCCCGATGATTGATT 277
               :::|||:::|||   ::::::   :::
     212 .......AsnLeuIleGlyGlyLysValLysAsnGlu............. 221

     278 TTTCTGTGGTATCGCGTAACGGC......GTGGCGGCATTGGCGGGCGAT 321
           :::|||:::|||   ||||||      ::::::   ||||||||||:::
     222 ..GlyValIleSerValAsnGlyGlySerIleSerLeuLeuAlaGlyGln 237

     322 CAA.....................................TATATTGTGAG 335
         :::                                     |||   :::::
     238 LysIleThrIleSerAspIleIleAsnProThrIleThrTyrSerIleAl 254

     336 CGTGGCACATAACGGCGGCTATAACAATGTTGATTTTGGTGCGGAGGGAA 385
         ::::   :::|||   :::   |||      |||:::   |||:::|||:
     254 aAlaProGluAsnGluAlaValAsnLeuGlyAspIlePheAlaLysGlyG 271

     386 GCAAT.......................CCCGATCAGCACCGCTTTTCT 411
         ::|||                       :::|||   :::::::|||
     271 lyAsnIleAsnValArgAlaAlaThrIleArgAsnGlnGlyLysLeuSer 287

     412 TACCAAATTGTGAAAAGAAATAATTATAAAGCAGGGACTAACGGCCATCC 461
           :::   |||:::::::::      |||:::|||
     288 AlaAspSerValSerLysAsp......LysSerGly.............. 297

     462 TTATGGCGGCGATTATCATATGCCGCGTTTGCACAAATTTGTCACAGATG 511
                                             :::::::|||   :::|
     298 ................................AsnIleValLeuSerA 303

     512 CAGAACCTGTTGAG.................ATGACCAGTTATATGGAT 543
         ||:::      |||                 :::::::::      :::
     303 laLysGluGlyGluAlaGluIleGlyGlyValIleSerAlaGlnAsnGln 319

     544 GGGTGGAAATACGCTGATTTAAATAAATACCCTGATCGTGTT......CG 587
               |||   :::   |||            |||:::|||      ::
     320 GlnAlaLysGlyGlyLysLeuMetIleThrGlyAspLysValThrLeuLy 336

     588 AATCGGAGCA.........................GGCAGACAATATT 610
         :   ||||||                        |||:::   |||
     336 sThrGlyAlaValIleAspLeuSerGlyLysGluGlyGlyGluThrTyrL 353

     611 GGCGGTCTGATGAA...GACGAACCCAATAAC.................. 639
               :::|||||||      |||   :::|||
     353 euGlyGlyAspGluArgGlyGluGlyLysAsnGlyIleGlnLeuAlaLys 369

     640 ..............CGCGAAAGTTCATATCATATTGCAAGC........ 666
                        :::   |||:::   ::::::::::::
     370 LysThrSerLeuGluLysGlySerThrIleAsnValSerGlyLysGluLy 386

     667 ..........GCATATTCTTGG............CTCGTCGGTGGCAATA 694
                    |||      |||            |||:::   ||||||
     386 sGlyGlyArgAlaIleValTrpGlyAspIleAlaLeuIleAspGlyAsnI 403
```

```
 695 CCTTTGCACAAAATGGATCAGGTGGTGGCACAGTCAACTTAGGTAGCGAA 744
 403 leAsnAlaGln...GlySerGly..................................... 409

 745 AAAATTAAACATAGCCCATATGGTTTTTTACCAACAGGAGGCTCATTTGG 794
 410 .....AspIleAlaLysThrGlyGlyPheValGl 419

 795 CGACAGTGGCTCACCAATGTTTATCTATGATGCCCAAAAGCAAAAGTGGT 844
 419 uThrSerGlyHisAspLeuPheIleLysAsp..... 429

 845 TAATTAATGGGGTATTGCAAACAGGCAACCCCTATATAGGAAAAAGCAAT 894
 430 .....AsnAlaIleValAspAla..... 435

 895 GGCTTCCAGCTAGTTCGTAAAGATTGGTTCTATGATGAAATCTTTGCTGG 944
 436 .....LysGluTrpLeuLeuAsp.....Ph 442

 945 AGATACCCATTCAGTATTCTACGAACCACATCAAAATGGGAAATACTTTT 994
 442 eAspAsnValSerIleAsnAlaGluThrAlaGlyArgSerAsnThrSerG 459

 995 TTAACGACAATAATAATGGCGCAGGAAAAATCGATGCCAAACATAAACAC 1044
 459 luAspAspGluTyrThrGlySerGly.....AsnSer 469

1045 TATTCTCTACCTTATAGATTAAAAACACGAACCGTTCAATTGTTTAATGT 1094
 470 AlaSerThrProLysArgAsnLysGluLysThrThr...LeuThrAsnTh 485

1095 TTCTTTATCCGAG.....ACAGCAA 1114
 485 rThrLeuGluSerIleLeuLysLysGlyThrPheValAsnIleThrAlaA 502

1115 GAGAACCTGTTTAT..... 1128
 502 snGlnArgIleTyrValAsnSerSerIleAsnLeuSerAsnGlySerLeu 518

1129 .....CATGCTGCAGGTGGGGTCAACAGTTATCGACC 1160
 519 ThrLeuTrpSerGluGlyArgSerGlyGlyGlyVal..... 530

1161 CAGACTGAATAAT.....GGAGAAAATA 1183
 531 .GluIleAsnAsnAspIleThrThrGlyAspAspThrArgGlyAlaAsnL 547

1184 TTTCCTTTATTGACAAAGGAAAAGGTGAATTGATACTTACCAGCAACATC 1233
 547 euThrIleTyrSerGlyGlyTrpValAsp......ValHisLysAsnIle 561

1234 AACCAAGGCGCGGGCGGTTTGTATTTTGAGGGTAATTTTACGGTCTCGCC 1283
 562 SerLeuGlyAla.....GlnGlyAsnIleAsnIleThrAl 573

1284 TAAA.....AACAACGAAACGTGGCAAGGCG 1309
 573 aLysGlnAspIleAlaPheGluLysGlySerAsnGlnValIleThrGlyG 590

1310 CGGGCGTTCATATCAGTGATGGCAGTACCGTTACTTGGAAAGTAAACGGC 1359
 590 lnGlyThr...IleThrSerGlyAsnGlnLysGlyPheArgPheAsnAsn 605

1360 GTGGCAAACGACCGCCTGTCCAAAATCGGCAAAGGCACGCTGCTGGTTCA 1409
 606 ValSer.....LeuAsnGlyThrGlySerGly...LeuGlnPheTh 618

1410 AGCCAAAGGGGAAAAC.....CAAGGCTCGG 1435
 618 rThrLysArgThrAsnLysTyrAlaIleThrAsnLysPheGluGlyThrL 635

1436 TCAGCGTGGGCGACGGTAAAGTCATCTTA.....GAT 1467
 635 euAsnIle...SerGlyLysValAsnIleSerMetValLeuProLysAsn 650

1468 CAGCAGGCGGACGATCAAGGCAAAAAACAAGCCTTT..... 1503
 651 GluSerGlyTyrAspLysPheLysGlyArgThrTyrTrpAsnLeuThrSe 667

1504 .....AGTGAAATCGGC.....TTGGTCAGCGGCAGGG 1531
 667 rLeuAsnValSerGluSerGlyGluPheAsnLeuThrIleAspSerArgG 684

1532 GGACG.....GTG 1539
 684 lySerAspSerAlaGlyThrLeuThrGlnProTyrAsnLeuAsnGlyIle 700

1540 CAACTGAATGCCGATAATCAGTTCAACCCCGACAAA..... 1575
 701 SerPheAsnLysAspThrThrPheAsnValGluArgAsnAlaArgValAs 717

1576 .....CTCT 1579
 717 nPheAspIleLysAlaProIleGlyIleAsnLysTyrSerSerLeuAsnT 734

1580 ATTTCGGCTTTCGCGGCGGACGTTTGGATTTGAACGGGCATTCGCTTTCG 1629
 734 yrAlaSerPheAsnGlyAsnIleSerValSerGlyGlyGlySerValAsp 750

1630 TTCCACCGCATT.....CAAAATACCGATGAAGGGGCGATGATTGT 1670
 751 PheThrLeuLeuAlaSerSerSerAsnValGlnThrProGlyValVall 767

1671 CAACCACAAT.....CAAGACAAAG 1690
 767 eAsnSerLysTyrPheAsnValSerThrGlySerSerLeuArgPheLysT 784

1691 AATCCACCGTTACCATTACAGGC.....AATAAAGATATTACT... 1728
 784 hrSerGlySerThrLysThrGlyPheSerIleGluLysAspLeuThrLeu 800

1729 ...ACAACCGGCAATAAC..... 1743
 801 AsnAlaThrGlyGlyAsnIleThrLeuLeuGlnValGluGlyThrAspGl 817

1744 .....AACAACTTGGATAGCAAAAAAGAAATTGCCTACAACGGT. 1782
 817 yMetIleGlyLysGlyIleValAlaLysLysAsnIleThrPheGluGlyG 834

1783 .....TGGTTTGGCGAGAAAGATGCA...ACCAAAACGAACGGGCGG 1821
 834 lyAsnIleThrPheGlySerArgLysAlaValThrGluIleGluGlyAsn 850

1822 CTCAATCTGAATTACCAACCGGAAGAAGCGGATCGCACTTTA..... 1863
 851 ValThrIleAsn.....AsnAsnAlaAsnValThrLeuIleGlySe 864

1863 ..... 1863
 864 rAspPheAspAsnHisGlnLysProLeuThrIleLysLysAspValIleI 881

1864 .....CTGCTTTCCGGCGGA......ACAAATTTAAACGGC 1893
 881 leAsnSerGlyAsnLeuThrAlaGlyGlyAsnIleValAsnIleAlaGly 897

1894 AATATCACG...CAAACAAACGGCAAACTGTTTTTCAGCGGCA..... 1933
 898 AsnLeuThrValGluSerAsn.AlaAsn...PheLysAlaIleThrAsnP 913

1934 .....GACCGACACCGCACGCCTACAATC 1957
 913 heThrPheAsnValGlyGlyLeuPheAspAsnLysGlyAsnSerAsnIle 929

1958 ATTTAGGAAGCGGGTGGTCAAAAATGGAAGGTATCCCACAAGGAGAAATC 2007
```

```
                 ||||||   :::::::|||          :::::::|||::
 930 SerIleAlaLysGlyGlyAlaArgPheLysAspIleAspAsnSerLysAs 946

2008 GTGTGGGACAACGATTGGATCGACCGC.ACATTTAAAGCGGAAAACTTCC 2056
     :   :::   |||   :::|||:::   |||::::::
 946 nLeuSerIleThrThrAsnSerSerSerThrTyrArg..............  958

2057 ATATTCAGGGCGGACAAGCGGTGGTTTCCCGCAATGTTGCCAAAGTGGAA 2106
                    ::::::::::|||   |||:::::::::   :::
 959 .................ThrIleIleSerGlyAsnIleThrAsnLysAsn 969

2107 GGCGATTGGCATTTAAGCAATCACGCC............CAAGCAGTTTT 2144
     ||||||   ::::::::::|||::::::            :::   :::
 970 GlyAspLeuAsnIleThrAsnGluGlySerAspThrGluMetGlnIleGl 986

2145 CGGTGTCGCACCGCATCAAAGCCACACAATCTGTACACGTTCGGACTGGA 2194
      |||   :::   :::::::::   ::::::          ||||||
 986 yGlyAspValSerGlnLysGluGlyAsnLeuThrIleSerSerAspLysI 1003

2195 CGGGTCTGACAAGTTGTACCGAAAAAACCATTACC...GACGATAAAGTG 2241
       :::::::|||            |||   ||||||          |||
1003 leAsnIleThr............LysGlnIleThrIleLysAlaGlyVal 1015

2242 ATTGCTTCATTGAGCAAGACCGACATCAGAGGCAATGTCAGCCTTGCCGA 2291
       ::::::   |||   :::|||      :::|||:::::::|||:::
1016 AspGlyGluAsnSerAspSerAspAlaThrAsnAsnAlaAsnLeuThrIl 1032

2292 TCACGCT.....................CATTTAAATCTCACAGGACTT. 2319
        :::                      ||||||:::::::|||:::
1032 eLysThrLysGluLeuLysLeuThrGlnAspLeuAsnIleSerGlyPheA 1049

2320 ................................GCCACACTCAACGGC 2334
                                     |||::::::::::
1049 snLysAlaGluIleThrAlaLysAspGlySerAspLeuThrIleGlyAsn 1065

2335 AATCTTAGTGCAGGCGGAGACACGCACTATACGGTTACGCGCAACGCCAC 2384
     :::   ||||||   |||   :::          ||||||   |||   ::
1066 ThrAsnSerAlaAspGlyThrAsnAlaLysLysValThrPheAsnGlnVa 1082

2385 CCAAAACGGCAACCTCAGCCTCGTGGGCAATGCCCAAGCAACATTTAATC 2434
     ::::::::::::::|||      |||:::      :::      ::::
1082 lLysAspSerLysIleSerAlaAspGlyHisLysValThrLeuHisSerL 1099

2435 AAGCCACATTAAACGGCAACACATCGGCTTCGGACAATGCTTCATTTAAT 2484
     :::::      :::|||:::::::::   :::::::::::::|||   |||
1099 ysValGluThrSerGlySerAsnAsnAsnThrGluAspSerSerAspAsn 1115

2485 CTAAGCAACAACGCCGTACAAAACGGCAGTCTGACGCTTTCCGACAACGC 2534
        ::::::   :::::::::      ::::::|||::::::::::|||
1116 AsnAlaGlyLeuThrIleAspAlaLysAsnValThrValAsnAsnAsnIl 1132

2535 TAAGGCAAAC......GTAAGCCATTCCGCACTCAACGGCAATGTC.... 2574
          ::::::      ||||||   ||||||   :::|||::::::
1132 eThrSerHisLysAlaValSerIleSerAlaThrSerGlyGluIleThrT 1149

2575 ...........TCCCTAGCCGATAAGGCAGTATTCCATTTTGAAAACAGC 2613
                 :::      :::   ::::::   :::   |||   :::
1149 hrLysThrGlyThrThrIleAsnAlaThrThrGlyAsnValGluIleThr 1165

2614 CGCTTTACCGGAAAAATCAGCGGCGGCAAGGATACG.........GCATT 2654
           ||||||:::|||   ||||||   ::::::          :::::
1166 AlaGlnThrGlySerIleLeuGlyGlyIleGluSerSerSerGlySerVa 1182

2655 ACACTTAAAAGACAGCGAATGGACGCTGCCGTCGGGCACGGAATTAGGCA 2704
     :   |||      :::|||   :::|||                :::::|
1182 lThrLeuThrAlaThrGluGlyAlaLeu............AlaValSerA 1195

2705 ATTTAAACCTTGACAACGCCACCATTACACTCAATTCCGCCTATCGACAC 2754
     ||::::::   :::::::::|||:::|||   ||||||
1195 snIleSerGlyAsnThrValThrValThrAlaAsnSer............ 1207

2755 GATGCGGCAGGCGCGCAAACCGGCAGTGCGGCAGATGCGCCGCGCCGCCG 2804
               ||||||   |||      |||::::::
```

```
1208 .........GlyAlaLeuThrThrLeuAlaGlySer.............. 1216

2805 TTCGCGCCGTTCCCTATTATCCGTTACGCCGCCAACTTCGGCAGAATCCC 2854
      :::   ::::::      |||||||||      ||||||:::::::|||
1217 .ThrIleLysGlyThrGluSerValThr......ThrSerSerGlnSerG 1231

2855 GTTTCAACACGCTGACGGTAAACGGCAAATTGAACGGTCAGGGAACA... 2901
                         ::::::|||   ::::::|||   ||||||
1231 ly............AspIleGlyGlyThrIleSerGly...GlyThrVal 1242

2902 ...TTCCGCTTTATGTCGGAACTCTTCGGCTACCGCAGCGGCAAATTGAA 2948
        :::       :::::::|||            ::::::|||:::||
1243 GluValLysAlaThrGluSerLeuThrThrGlnSerAsnSerLysIleLy 1259

2949 GCTGGCGGAAAGTTCCGAAGGCACTTACACCTTGGCTGTC.......... 2988
     |   |||   :::::::|||:::::::   |||   |||:::
1259 s...AlaThrThrGlyGluAlaAsnValThrSerAlaThrGlyThrIleG 1275

2989 ..............AACAATACCGGCAACGAACCCGTAAGTCTCGAGCAA 3024
                     :::|||||||   |||      ::::::      :::
1275 lyGlyThrIleSerGlyAsnThrValAsnValThrAlaAsnAlaGlyAsp 1291

3025 TTGACGGTA.......................GTGGAAGGAAAAGACAA 3050
     |||||||||                      :::|||||||      ::
1292 LeuThrValGlyAsnGlyAlaGluIleAsnAlaThrGluGlyAlaAlaTh 1308

3051 CACACCGCTGTCCGAAAATCTTAATTTCACCCTGCAAAACGAACACGTCG 3100
     :          |||::::::          |||      ::::::|||:::
1308 rLeuThrThrSerSerGlyLysLeuThrThrGluAlaSerSerHisIleT 1325

3101 ATGCCGGCGCATGGCGTTATCAGCTTATCCGCAAAGACGGCGAGTTCCGC 3150
       ::: |||   |||   ::::::||||||||||:::   :::|||   ||
1325 hrSer.AlaLysGlyGlnValAsnLeuSerAlaGlnAspGlySerValAl 1341

3151 CTGCATAATCCGGTCAAAGAACAAGAGCTTTCCGACA 3187
     |      |||::::::   |||   ::::::   |||
1341 aGlySerIleAsnAlaAlaAsnValThrLeuAsnThr 1353

seq_name: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:US-08-469-880-2

seq_documentation_block:
; Sequence 2, Application US/08469880
; Patent No. 5876733
;  GENERAL INFORMATION:
;    APPLICANT:  Barenkamp, Stephen J.
;    TITLE OF INVENTION:  High Molecular Weight Surface Proteins
;    TITLE OF INVENTION:  of No. 5876733-Typeable Haemophilus
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Shoemaker and Mattare, Ltd.
;      STREET:  2001 Jefferson Davis Hwy., 1203 Crystal Plaza
;      STREET:  Bldg. 1
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22202-0286
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/469,880
;      FILING DATE:  06-JUN-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9205704.1
;      FILING DATE:  16-MAR-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US PCT/US93/02166
;      FILING DATE:  16-MAR-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/302,832
;
```

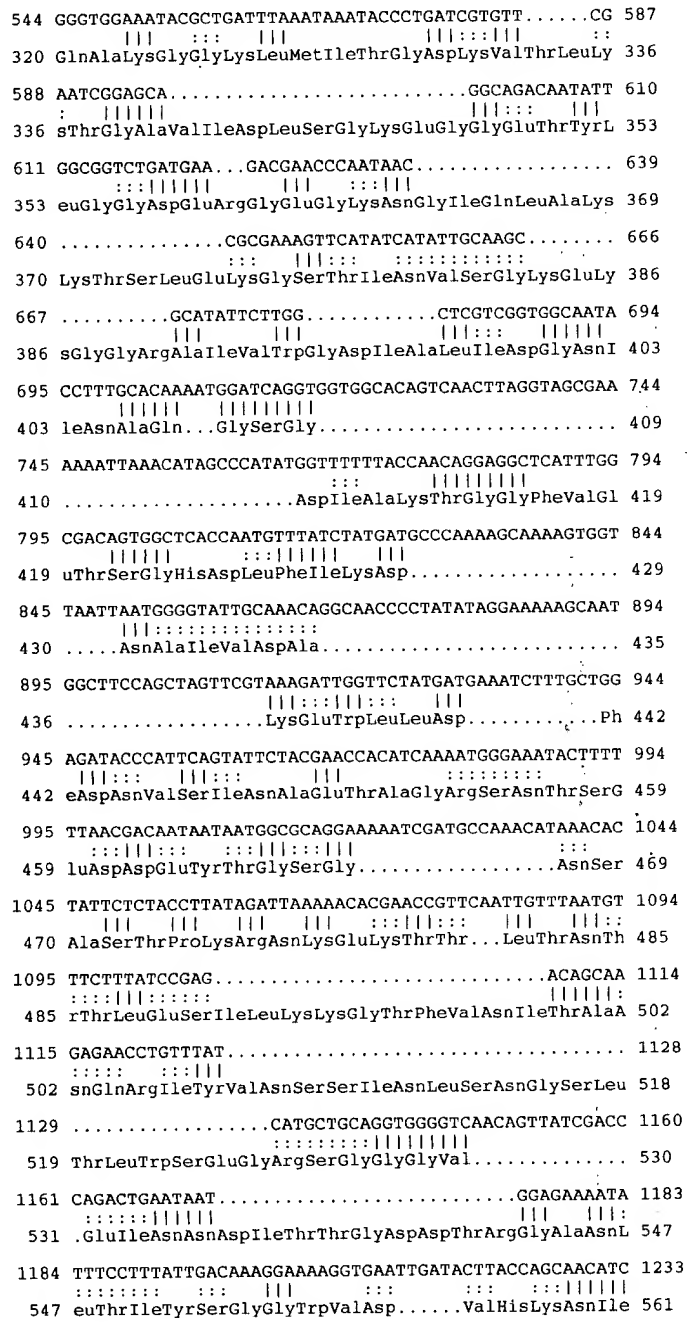
```
;       FILING DATE:  16-SEP-1994
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Berkstresser, Jerry W
;       REGISTRATION NUMBER:  22,651
;       REFERENCE/DOCKET NUMBER:  1038-516 MIS:vg
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (703) 415-0810
;       TELEFAX:  (703) 415-0813
;   INFORMATION FOR SEQ ID NO:  2:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1536 amino acids
;       TYPE:  amino acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA (genomic)
US-08-469-880-2


alignment_scores:
            Quality:  265.00              Length:   1297
              Ratio:  0.438                Gaps:     73
 Percent Similarity:  46.646   Percent Identity: 21.126

alignment_block:
 US-09-303-518D-653 x US-08-469-880-2   ..

 Align seg 1/1  to: US-08-469-880-2  from: 1  to: 1536

     178 AATAAAGGCAAGTTTGCAGTCGGGGCGAAAGATATTGAGGTTTACAACAA 227
         |||   |||   ::::::||||||   ||||||   :::|||
     199 AsnHisGlyLeuIleThrValGly...LysAspGlySerVal........ 211

     228 AAAAGGGGAGTTGGTCGGCAAATCGATGACGAAAGCCCCGATGATTGATT 277
             :::|||:::|||   ::::::   :::
     212 .......AsnLeuIleGlyGlyLysValLysAsnGlu.............. 221

     278 TTTCTGTGGTATCGCGTAACGGC......GTGGCGGCATTGGCGGGCGAT 321
           :::|||:::|||   ||||||      ::::::   |||||||||:::
     222 ..GlyValIleSerValAsnGlyGlySerIleSerLeuLeuAlaGlyGln 237

     322 CAA.....................................TATATTGTGAG 335
         :::                                     |||   :::::
     238 LysIleThrIleSerAspIleIleAsnProThrIleThrTyrSerIleAl 254

     336 CGTGGCACATAACGGCGGCTATAACAATGTTGATTTTGGTGCGGAGGGAA 385
         ::::   :::|||   :::   |||      |||:::   |||:::|||:
     254 aAlaProGluAsnGluAlaValAsnLeuGlyAspIlePheAlaLysGlyG 271

     386 GCAAT.......................CCCGATCAGCACCGCTTTTCT 411
         ::|||                       :::|||   ::::::|||
     271 lyAsnIleAsnValArgAlaAlaThrIleArgAsnGlnGlyLysLeuSer 287

     412 TACCAAATTGTGAAAAGAAATAATTATAAAGCAGGGACTAACGGCCATCC 461
            :::   |||:::::::::      |||:::|||
     288 AlaAspSerValSerLysAsp......LysSerGly.............. 297

     462 TTATGGCGGCGATTATCATATGCCGCGTTTGCACAAATTTGTCACAGATG 511
                                     :::::::|||   :::|
     298 ................................AsnIleValLeuSerA 303

     512 CAGAACCTGTTGAG.................ATGACCAGTTATATGGAT 543
         ||:::      |||                 :::::::::      :::
     303 laLysGluGlyGluAlaGluIleGlyGlyValIleSerAlaGlnAsnGln 319

     544 GGGTGGAAATACGCTGATTTAAATAAATACCCTGATCGTGTT......CG 587
             |||   :::   |||           |||:::|||      ::
     320 GlnAlaLysGlyGlyLysLeuMetIleThrGlyAspLysValThrLeuLy 336

     588 AATCGGAGCA.........................GGCAGACAATATT 610
         :   ||||||                        |||:::   |||
     336 sThrGlyAlaValIleAspLeuSerGlyLysGluGlyGlyGluThrTyrL 353

     611 GGCGGTCTGATGAA...GACGAACCCAATAAC.................. 639

           :::|||||||      |||   :::|||
     353 euGlyGlyAspGluArgGlyGluGlyLysAsnGlyIleGlnLeuAlaLys 369

     640 ..............CGCGAAAGTTCATATCATATTGCAAGC........ 666
                      :::   |||:::   ::::::::::::
     370 LysThrSerLeuGluLysGlySerThrIleAsnValSerGlyLysGluLy 386

     667 ..........GCATATTCTTGG............CTCGTCGGTGGCAATA 694
                 |||      |||            |||:::   ||||||
     386 sGlyGlyArgAlaIleValTrpGlyAspIleAlaLeuIleAspGlyAsnI 403

     695 CCTTTGCACAAAATGGATCAGGTGGTGGCACAGTCAACTTAGGTAGCGAA 744
         ||||||   |||||||||
     403 leAsnAlaGln...GlySerGly........................... 409

     745 AAAATTAAACATAGCCCATATGGTTTTTTACCAACAGGAGGCTCATTTGG 794
                                  :::      |||||||||
     410 .....................AspIleAlaLysThrGlyGlyPheValGl 419

     795 CGACAGTGGCTCACCAATGTTTATCTATGATGCCCAAAAGCAAAAGTGGT 844
         ||||||      :::||||||   |||
     419 uThrSerGlyHisAspLeuPheIleLysAsp.................... 429

     845 TAATTAATGGGGTATTGCAAACAGGCAACCCCTATATAGGAAAAAGCAAT 894
         |||:::::::::::::::::
     430 .....AsnAlaIleValAspAla............................ 435

     895 GGCTTCCAGCTAGTTCGTAAAGATTGGTTCTATGATGAAATCTTTGCTGG 944
                            |||:::|||:::   |||
     436 ..................LysGluTrpLeuLeuAsp.............Ph 442

     945 AGATACCCATTCAGTATTCTACGAACCACATCAAAATGGGAAATACTTTT 994
          |||:::   |||:::      |||          :::::::::
     442 eAspAsnValSerIleAsnAlaGluThrAlaGlyArgSerAsnThrSerG 459

     995 TTAACGACAATAATAATGGCGCAGGAAAAATCGATGCCAAACATAAACAC 1044
           :::|||:::   :::|||:::|||                     :::
     459 luAspAspGluTyrThrGlySerGly.................AsnSer 469

    1045 TATTCTCTACCTTATAGATTAAAAACACGAACCGTTCAATTGTTTAATGT 1094
         |||   |||   |||   |||   :::|||:::   |||   |||::
     470 AlaSerThrProLysArgAsnLysGluLysThrThr...LeuThrAsnTh 485

    1095 TTCTTTATCCGAG.............................ACAGCAA 1114
         ::::|||:::::                              ||||||:
     485 rThrLeuGluSerIleLeuLysLysGlyThrPheValAsnIleThrAlaA 502

    1115 GAGAACCTGTTTAT................................. 1128
         :::::   :::|||
     502 snGlnArgIleTyrValAsnSerSerIleAsnLeuSerAsnGlySerLeu 518

    1129 ..................CATGCTGCAGGTGGGGTCAACAGTTATCGACC 1160
                            :::::::::|||||||||
     519 ThrLeuTrpSerGluGlyArgSerGlyGlyGlyVal.............. 530

    1161 CAGACTGAATAAT..........................GGAGAAAATA 1183
         ::::::||||||                           |||   |||:
     531 .GluIleAsnAsnAspIleThrThrGlyAspAspThrArgGlyAlaAsnL 547

    1184 TTTCCTTTATTGACAAAGGAAAAGGTGAATTGATACTTACCAGCAACATC 1233
         ::::::::   :::   |||      :::      :::   :::||||||
     547 euThrIleTyrSerGlyGlyTrpValAsp......ValHisLysAsnIle 561

    1234 AACCAAGGCGCGGGCGGTTTGTATTTTGAGGGTAATTTTACGGTCTCGCC 1283
         :::   ||||||              :::||||||:::::::::::::
     562 SerLeuGlyAla..............GlnGlyAsnIleAsnIleThrAl 573

    1284 TAAA........................AACAACGAAACGTGGCAAGGCG 1309
          |||                       :::|||::::::      |||
     573 aLysGlnAspIleAlaPheGluLysGlySerAsnGlnValIleThrGlyG 590

    1310 CGGGCGTTCATATCAGTGATGGCAGTACCGTTACTTGGAAAGTAAACGGC 1359
         |||:::   |||::::::|||:::         ::::::   |||:::
```

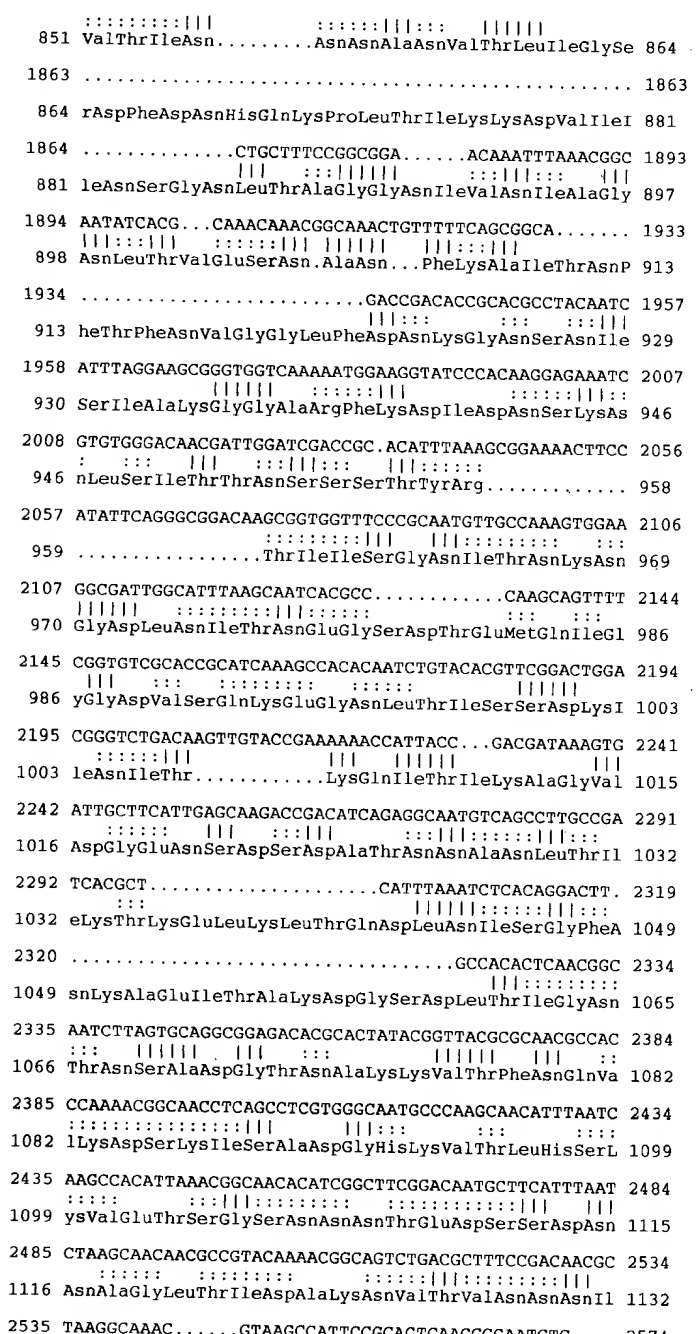

590 lnGlyThr...IleThrSerGlyAsnGlnLysGlyPheArgPheAsnAsn 605
1360 GTGGCAAACGACCGCCTGTCCAAAATCGGCAAAGGCACGCTGCTGGTTCA 1409
606 ValSer.....LeuAsnGlyThrGlySerGly...LeuGlnPheTh 618
1410 AGCCAAAGGGGAAAAC.....CAAGGCTCGG 1435
618 rThrLysArgThrAsnLysTyrAlaIleThrAsnLysPheGluGlyThrL 635
1436 TCAGCGTGGGCGACGGTAAAGTCATCTTA.....GAT 1467
635 euAsnIle...SerGlyLysValAsnIleSerMetValLeuProLysAsn 650
1468 CAGCAGGCGGACGATCAAGGCAAAAAACAAGCCTTT..... 1503
651 GluSerGlyTyrAspLysPheLysGlyArgThrTyrTrpAsnLeuThrSe 667
1504 .....AGTGAAATCGGC.....TTGGTCAGCGGCAGGG 1531
667 rLeuAsnValSerGluSerGlyGluPheAsnLeuThrIleAspSerArgG 684
1532 GGACG.....GTG 1539
684 lySerAspSerAlaGlyThrLeuThrGlnProTyrAsnLeuAsnGlyIle 700
1540 CAACTGAATGCCGATAATCAGTTCAACCCCGACAAA..... 1575
701 SerPheAsnLysAspThrThrPheAsnValGluArgAsnAlaArgValAs 717
1576 .....CTCT 1579
717 nPheAspIleLysAlaProIleGlyIleAsnLysTyrSerSerLeuAsnT 734
1580 ATTTCGGCTTTCGCGGCGGACGTTTGGATTTGAACGGGCATTCGCTTTCG 1629
734 yrAlaSerPheAsnGlyAsnIleSerValSerGlyGlyGlySerValAsp 750
1630 TTCCACCGCATT.....CAAAATACCGATGAAGGGGCGATGATTGT 1670
751 PheThrLeuLeuAlaSerSerSerAsnValGlnThrProGlyValVall 767
1671 CAACCACAAT.....CAAGACAAAG 1690
767 eAsnSerLysTyrPheAsnValSerThrGlySerSerLeuArgPheLysT 784
1691 AATCCACCGTTACCATTACAGGC.....AATAAAGATATTACT... 1728
784 hrSerGlySerThrLysThrGlyPheSerIleGluLysAspLeuThrLeu 800
1729 ...ACAACCGGCAATAAC..... 1743
801 AsnAlaThrGlyGlyAsnIleThrLeuLeuGlnValGluGlyThrAspGl 817
1744 .....AACAACTTGGATAGCAAAAAAGAAATTGCCTACAACGGT. 1782
817 yMetIleGlyLysGlyIleValAlaLysLysAsnIleThrPheGluGlyG 834
1783 .....TGGTTTGGCGAGAAAGATGCA...ACCAAAACGAACGGGCGG 1821
834 lyAsnIleThrPheGlySerArgLysAlaValThrGluIleGluGlyAsn 850
1822 CTCAATCTGAATTACCAACCGGAAGAAGCGGATCGCACTTTA..... 1863
851 ValThrIleAsn.....AsnAsnAlaAsnValThrLeuIleGlySe 864
1863 1863
864 rAspPheAspAsnHisGlnLysProLeuThrIleLysLysAspValIleI 881
1864 .....CTGCTTTCCGGCGGA.....ACAAATTTAAACGGC 1893
881 leAsnSerGlyAsnLeuThrAlaGlyGlyAsnIleValAsnIleAlaGly 897
1894 AATATCACG...CAAACAAACGGCAAACTGTTTTTCAGCGGCA..... 1933
898 AsnLeuThrValGluSerAsn.AlaAsn...PheLysAlaIleThrAsnP 913
1934 .....GACCGACACCGCACGCCTACAATC 1957
913 heThrPheAsnValGlyGlyLeuPheAspAsnLysGlyAsnSerAsnIle 929
1958 ATTTAGGAAGCGGGTGGTCAAAAATGGAAGGTATCCCACAAGGAGAAATC 2007
930 SerIleAlaLysGlyGlyAlaArgPheLysAspIleAspAsnSerLysAs 946
2008 GTGTGGGACAACGATTGGATCGACCGC.ACATTTAAAGCGGAAAACTTCC 2056
946 nLeuSerIleThrThrAsnSerSerSerThrTyrArg..... 958
2057 ATATTCAGGGCGGACAAGCGGTGGTTTCCCGCAATGTTGCCAAAGTGGAA 2106
959 .....ThrIleIleSerGlyAsnIleThrAsnLysAsn 969
2107 GGCGATTGGCATTTAAGCAATCACGCC.....CAAGCAGTTTT 2144
970 GlyAspLeuAsnIleThrAsnGluGlySerAspThrGluMetGlnIleGl 986
2145 CGGTGTCGCACCGCATCAAAGCCACACAATCTGTACACGTTCGGACTGGA 2194
986 yGlyAspValSerGlnLysGluGlyAsnLeuThrIleSerSerAspLysI 1003
2195 CGGGTCTGACAAGTTGTACCGAAAAAACCATTACC...GACGATAAAGTG 2241
1003 leAsnIleThr.....LysGlnIleThrIleLysAlaGlyVal 1015
2242 ATTGCTTCATTGAGCAAGACCGACATCAGAGGCAATGTCAGCCTTGCCGA 2291
1016 AspGlyGluAsnSerAspSerAspAlaThrAsnAsnAlaAsnLeuThrIl 1032
2292 TCACGCT.....CATTTAAATCTCACAGGACTT. 2319
1032 eLysThrLysGluLeuLysLeuThrGlnAspLeuAsnIleSerGlyPheA 1049
2320 .....GCCACACTCAACGGC 2334
1049 snLysAlaGluIleThrAlaLysAspGlySerAspLeuThrIleGlyAsn 1065
2335 AATCTTAGTGCAGGCGGAGACACGCACTATACGGTTACGCGCAACGCCAC 2384
1066 ThrAsnSerAlaAspGlyThrAsnAlaLysLysValThrPheAsnGlnVa 1082
2385 CCAAAACGGCAACCTCAGCCTCGTGGGCAATGCCCAAGCAACATTTAATC 2434
1082 lLysAspSerLysIleSerAlaAspGlyHisLysValThrLeuHisSerL 1099
2435 AAGCCACATTAAACGGCAACACATCGGCTTCGGACAATGCTTCATTTAAT 2484
1099 ysValGluThrSerGlySerAsnAsnAsnThrGluAspSerSerAspAsn 1115
2485 CTAAGCAACAACGCCGTACAAAACGGCAGTCTGACGCTTTCCGACAACGC 2534
1116 AsnAlaGlyLeuThrIleAspAlaLysAsnValThrValAsnAsnAsnIl 1132
2535 TAAGGCAAAC.....GTAAGCCATTCCGCACTCAACGGCAATGTC.... 2574
1132 eThrSerHisLysAlaValSerIleSerAlaThrSerGlyGluIleThrT 1149
2575 .....TCCCTAGCCGATAAGGCAGTATTCCATTTTGAAAACAGC 2613
1149 hrLysThrGlyThrThrIleAsnAlaThrThrGlyAsnValGluIleThr 1165
2614 CGCTTTACCGGAAAAATCAGCGGCGGCAAGGATACG.....GCATT 2654
1166 AlaGlnThrGlySerIleLeuGlyGlyIleGluSerSerSerGlySerVa 1182

```
 2655 ACACTTAAAAGACAGCGAATGGACGCTGCCGTCGGGCACGGAATTAGGCA 2704
      :  |||      :::|||  :::|||                  ::::::|
 1182 lThrLeuThrAlaThrGluGlyAlaLeu.............AlaValSerA 1195

 2705 ATTTAAACCTTGACAACGCCACCATTACACTCAATTCCGCCTATCGACAC 2754
      ||:::::::  :::::::::|||:::|||  ||||||
 1195 snIleSerGlyAsnThrValThrValThrAlaAsnSer............. 1207

 2755 GATGCGGCAGGCGCGCAAACCGGCAGTGCGGCAGATGCGCCGCGCCGCCG 2804
             ||||||  |||     |||::::::
 1208 .........GlyAlaLeuThrThrLeuAlaGlySer.............. 1216

 2805 TTCGCGCCGTTCCCTATTATCCGTTACGCCGCCAACTTCGGCAGAATCCC 2854
       :::  ::::::      |||||||||      ||||||:::::|||
 1217 .ThrIleLysGlyThrGluSerValThr......ThrSerSerGlnSerG 1231

 2855 GTTTCAACACGCTGACGGTAAACGGCAAATTGAACGGTCAGGGAACA... 2901
                    ::::::|||   :::::::|||   ||||||
 1231 ly.............AspIleGlyGlyThrIleSerGly...GlyThrVal 1242

 2902 ...TTCCGCTTTATGTCGGAACTCTTCGGCTACCGCAGCGGCAAATTGAA 2948
         :::      :::::::|||             ::::::|||:::||
 1243 GluValLysAlaThrGluSerLeuThrThrGlnSerAsnSerLysIleLy 1259

 2949 GCTGGCGGAAAGTTCCGAAGGCACTTACACCTTGGCTGTC.......... 2988
      |   |||   :::::::|||::::::   |||  |||:::
 1259 s...AlaThrThrGlyGluAlaAsnValThrSerAlaThrGlyThrIleG 1275

 2989 ..............AACAATACCGGCAACGAACCCGTAAGTCTCGAGCAA 3024
                    :::||||||   |||     ::::::      :::
 1275 lyGlyThrIleSerGlyAsnThrValAsnValThrAlaAsnAlaGlyAsp 1291

 3025 TTGACGGTA.....................GTGGAAGGAAAAGACAA 3050
      |||||||||                    :::|||||     ::
 1292 LeuThrValGlyAsnGlyAlaGluIleAsnAlaThrGluGlyAlaAlaTh 1308

 3051 CACACCGCTGTCCGAAAATCTTAATTTCACCCTGCAAAACGAACACGTCG 3100
      :         |||::::::         |||     ::::::|||:::
 1308 rLeuThrThrSerSerGlyLysLeuThrThrGluAlaSerSerHisIleT 1325

 3101 ATGCCGGCGCATGGCGTTATCAGCTTATCCGCAAAGACGGCGAGTTCCGC 3150
        ::: |||   |||   ::::::|||||||||::   :::|||   ||
 1325 hrSer.AlaLysGlyGlnValAsnLeuSerAlaGlnAspGlySerValAl 1341

 3151 CTGCATAATCCGGTCAAAGAACAAGAGCTTTCCGACA 3187
      |      |||::::::   |||   ::::::  |||
 1341 aGlySerIleAsnAlaAlaAsnValThrLeuAsnThr 1353
```

seq_name: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:US-08-728-470-2

```
seq_documentation_block:
; Sequence 2, Application US/08728470
; Patent No. 5928651
;  GENERAL INFORMATION:
;    APPLICANT:  Barenkamp, Stephen J
;    TITLE OF INVENTION:  High Molecular Weight Surface Proteins
;    TITLE OF INVENTION:  of No. 5928651-Typeable Haemophilus
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Shoemaker and Mattare, Ltd.
;      STREET:  2001 Jefferson Davis Hwy., 1203 Crystal Plaza
;      STREET:  Bldg. 1
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22202-0286
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/728,470
;      FILING DATE:
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/302,832
;      FILING DATE:  16-MAR-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US PCT/US93/02166
;      FILING DATE:  16-MAR-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9205704.1
;      FILING DATE:  16-MAR-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Berkstresser, Jerry W
;      REGISTRATION NUMBER:  22,651
;      REFERENCE/DOCKET NUMBER:  1038-633
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 415-0810
;      TELEFAX:  (703) 415-0813
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1536 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:
US-08-728-470-2


alignment_scores:
            Quality:  265.00              Length:   1297
              Ratio:  0.438                 Gaps:     73
 Percent Similarity:  46.646   Percent Identity: 21.126

alignment_block:
 US-09-303-518D-653 x US-08-728-470-2   ..

 Align seg 1/1  to: US-08-728-470-2  from: 1  to: 1536

     178 AATAAAGGCAAGTTTGCAGTCGGGGCGAAAGATATTGAGGTTTACAACAA 227
         |||   |||   ::::::||||||   ||||||   :::|||
     199 AsnHisGlyLeuIleThrValGly...LysAspGlySerVal........ 211

     228 AAAAGGGGAGTTGGTCGGCAAATCGATGACGAAAGCCCCGATGATTGATT 277
               :::|||:::|||   ::::::   :::
     212 .......AsnLeuIleGlyGlyLysValLysAsnGlu.............. 221

     278 TTTCTGTGGTATCGCGTAACGGC......GTGGCGGCATTGGCGGGCGAT 321
           :::|||:::|||   ||||||      ::::::   |||||||||:::
     222 ..GlyValIleSerValAsnGlyGlySerIleSerLeuLeuAlaGlyGln 237

     322 CAA.....................................TATATTGTGAG 335
         :::                                    |||   :::::
     238 LysIleThrIleSerAspIleIleAsnProThrIleThrTyrSerIleAl 254

     336 CGTGGCACATAACGGCGGCTATAACAATGTTGATTTTGGTGCGGAGGGAA 385
         ::::   :::|||   :::   |||      |||:::   |||:::|||:
     254 aAlaProGluAsnGluAlaValAsnLeuGlyAspIlePheAlaLysGlyG 271

     386 GCAAT.......................CCCGATCAGCACCGCTTTTCT 411
         ::|||                        :::|||   ::::::|||
     271 lyAsnIleAsnValArgAlaAlaThrIleArgAsnGlnGlyLysLeuSer 287

     412 TACCAAATTGTGAAAAGAAATAATTATAAAGCAGGGACTAACGGCCATCC 461
            :::   |||:::::::::      |||:::|||
     288 AlaAspSerValSerLysAsp......LysSerGly.............. 297

     462 TTATGGCGGCGATTATCATATGCCGCGTTTGCACAAATTTGTCACAGATG 511
                                      :::::::|||   :::|
     298 ................................AsnIleValLeuSerA 303

     512 CAGAACCTGTTGAG.................ATGACCAGTTATATGGAT 543
         ||:::     |||                 :::::::::      :::
     303 laLysGluGlyGluAlaGluIleGlyGlyValIleSerAlaGlnAsnGln 319
```


544 GGGTGGAAATACGCTGATTTAAATAAATACCCTGATCGTGTT.....CG 587
320 GlnAlaLysGlyGlyLysLeuMetIleThrGlyAspLysValThrLeuLy 336
588 AATCGGAGCA.....GGCAGACAATATT 610
336 sThrGlyAlaValIleAspLeuSerGlyLysGluGlyGlyGluThrTyrL 353
611 GGCGGTCTGATGAA...GACGAACCCAATAAC..... 639
353 euGlyGlyAspGluArgGlyGluGlyLysAsnGlyIleGlnLeuAlaLys 369
640 .....CGCGAAAGTTCATATCATATTGCAAGC..... 666
370 LysThrSerLeuGluLysGlySerThrIleAsnValSerGlyLysGluLy 386
667 .....GCATATTCTTGG.....CTCGTCGGTGGCAATA 694
386 sGlyGlyArgAlaIleValTrpGlyAspIleAlaLeuIleAspGlyAsnI 403
695 CCTTTGCACAAAATGGATCAGGTGGTGGCACAGTCAACTTAGGTAGCGAA 744
403 leAsnAlaGln...GlySerGly..... 409
745 AAAATTAAACATAGCCCATATGGTTTTTTACCAACAGGAGGCTCATTTGG 794
410 .....AspIleAlaLysThrGlyGlyPheValGl 419
795 CGACAGTGGCTCACCAATGTTTATCTATGATGCCCAAAAGCAAAAGTGGT 844
419 uThrSerGlyHisAspLeuPheIleLysAsp..... 429
845 TAATTAATGGGGTATTGCAAACAGGCAACCCCTATATAGGAAAAAGCAAT 894
430 .....AsnAlaIleValAspAla..... 435
895 GGCTTCCAGCTAGTTCGTAAAGATTGGTTCTATGATGAAATCTTTGCTGG 944
436 .....LysGluTrpLeuLeuAsp.....Ph 442
945 AGATACCCATTCAGTATTCTACGAACCACATCAAAATGGGAAATACTTTT 994
442 eAspAsnValSerIleAsnAlaGluThrAlaGlyArgSerAsnThrSerG 459
995 TTAACGACAATAATAATGGCGCAGGAAAAATCGATGCCAAACATAAACAC 1044
459 luAspAspGluTyrThrGlySerGly.....AsnSer 469
1045 TATTCTCTACCTTATAGATTAAAAACACGAACCGTTCAATTGTTTAATGT 1094
470 AlaSerThrProLysArgAsnLysGluLysThrThr...LeuThrAsnTh 485
1095 TTCTTTATCCGAG.....ACAGCAA 1114
485 rThrLeuGluSerIleLeuLysLysGlyThrPheValAsnIleThrAlaA 502
1115 GAGAACCTGTTTAT..... 1128
502 snGlnArgIleTyrValAsnSerSerIleAsnLeuSerAsnGlySerLeu 518
1129 .....CATGCTGCAGGTGGGGTCAACAGTTATCGACC 1160
519 ThrLeuTrpSerGluGlyArgSerGlyGlyGlyVal..... 530
1161 CAGACTGAATAAT.....GGAGAAAATA 1183
531 .GluIleAsnAsnAspIleThrThrGlyAspAspThrArgGlyAlaAsnL 547
1184 TTTCCTTTATTGACAAAGGAAAAGGTGAATTGATACTTACCAGCAACATC 1233
547 euThrIleTyrSerGlyGlyTrpValAsp.....ValHisLysAsnIle 561
1234 AACCAAGGCGCGGGCGGTTTGTATTTTGAGGGTAATTTTACGGTCTCGCC 1283
562 SerLeuGlyAla.....GlnGlyAsnIleAsnIleThrAl 573
1284 TAAA.....AACAACGAAACGTGGCAAGGCG 1309
573 aLysGlnAspIleAlaPheGluLysGlySerAsnGlnValIleThrGlyG 590
1310 CGGGCGTTCATATCAGTGATGGCAGTACCGTTACTTGGAAAGTAAACGGC 1359
590 lnGlyThr...IleThrSerGlyAsnGlnLysGlyPheArgPheAsnAsn 605
1360 GTGGCAAACGACCGCCTGTCCAAAATCGGCAAAGGCACGCTGCTGGTTCA 1409
606 ValSer.....LeuAsnGlyThrGlySerGly...LeuGlnPheTh 618
1410 AGCCAAAGGGGAAAAC.....CAAGGCTCGG 1435
618 rThrLysArgThrAsnLysTyrAlaIleThrAsnLysPheGluGlyThrL 635
1436 TCAGCGTGGGCGACGGTAAAGTCATCTTA.....GAT 1467
635 euAsnIle...SerGlyLysValAsnIleSerMetValLeuProLysAsn 650
1468 CAGCAGGCGGACGATCAAGGCAAAAAACAAGCCTTT..... 1503
651 GluSerGlyTyrAspLysPheLysGlyArgThrTyrTrpAsnLeuThrSe 667
1504 .....AGTGAAATCGGC.....TTGGTCAGCGGCAGGG 1531
667 rLeuAsnValSerGluSerGlyGluPheAsnLeuThrIleAspSerArgG 684
1532 GGACG.....GTG 1539
684 lySerAspSerAlaGlyThrLeuThrGlnProTyrAsnLeuAsnGlyIle 700
1540 CAACTGAATGCCGATAATCAGTTCAACCCCGACAAA..... 1575
701 SerPheAsnLysAspThrThrPheAsnValGluArgAsnAlaArgValAs 717
1576 .....CTCT 1579
717 nPheAspIleLysAlaProIleGlyIleAsnLysTyrSerSerLeuAsnT 734
1580 ATTTCGGCTTTCGCGGCGGACGTTTGGATTTGAACGGGCATTCGCTTTCG 1629
734 yrAlaSerPheAsnGlyAsnIleSerValSerGlyGlyGlySerValAsp 750
1630 TTCCACCGCATT.....CAAAATACCGATGAAGGGGCGATGATTGT 1670
751 PheThrLeuLeuAlaSerSerSerAsnValGlnThrProGlyValVallI 767
1671 CAACCACAAT.....CAAGACAAAG 1690
767 eAsnSerLysTyrPheAsnValSerThrGlySerSerLeuArgPheLysT 784
1691 AATCCACCGTTACCATTACAGGC.....AATAAAGATATTACT... 1728
784 hrSerGlySerThrLysThrGlyPheSerIleGluLysAspLeuThrLeu 800
1729 ...ACAACCGGCAATAAC..... 1743
801 AsnAlaThrGlyGlyAsnIleThrLeuLeuGlnValGluGlyThrAspGl 817
1744 .....AACAACTTGGATAGCAAAAAAGAAATTGCCTACAACGGT. 1782
817 yMetIleGlyLysGlyIleValAlaLysLysAsnIleThrPheGluGlyG 834
1783 .....TGGTTTGGCGAGAAAGATGCA...ACCAAAACGAACGGGCGG 1821
834 lyAsnIleThrPheGlySerArgLysAlaValThrGluIleGluGlyAsn 850
1822 CTCAATCTGAATTACCAACCGGAAGAAGCGGATCGCACTTTA..... 1863



```
851 ValThrIleAsn.....AsnAsnAlaAsnValThrLeuIleGlySe 864

1863 ..................................................... 1863

864 rAspPheAspAsnHisGlnLysProLeuThrIleLysLysAspValIleI 881

1864 .............CTGCTTTCCGGCGGA......ACAAATTTAAACGGC 1893
881 leAsnSerGlyAsnLeuThrAlaGlyGlyAsnIleValAsnIleAlaGly 897

1894 AATATCACG...CAAACAAACGGCAAACTGTTTTTCAGCGGCA..... 1933
898 AsnLeuThrValGluSerAsn.AlaAsn...PheLysAlaIleThrAsnP 913

1934 .........................GACCGACACCGCACGCCTACAATC 1957
913 heThrPheAsnValGlyGlyLeuPheAspAsnLysGlyAsnSerAsnIle 929

1958 ATTTAGGAAGCGGGTGGTCAAAAATGGAAGGTATCCCACAAGGAGAAATC 2007
930 SerIleAlaLysGlyGlyAlaArgPheLysAspIleAspAsnSerLysAs 946

2008 GTGTGGGACAACGATTGGATCGACCGC.ACATTTAAAGCGGAAAACTTCC 2056
946 nLeuSerIleThrThrAsnSerSerSerThrTyrArg..... 958

2057 ATATTCAGGGCGGACAAGCGGTGGTTTCCCGCAATGTTGCCAAAGTGGAA 2106
959 .................ThrIleIleSerGlyAsnIleThrAsnLysAsn 969

2107 GGCGATTGGCATTTAAGCAATCACGCC.....CAAGCAGTTTT 2144
970 GlyAspLeuAsnIleThrAsnGluGlySerAspThrGluMetGlnIleGl 986

2145 CGGTGTCGCACCGCATCAAAGCCACACAATCTGTACACGTTCGGACTGGA 2194
986 yGlyAspValSerGlnLysGluGlyAsnLeuThrIleSerSerAspLysI 1003

2195 CGGGTCTGACAAGTTGTACCGAAAAAACCATTACC...GACGATAAAGTG 2241
1003 leAsnIleThr.....LysGlnIleThrIleLysAlaGlyVal 1015

2242 ATTGCTTCATTGAGCAAGACCGACATCAGAGGCAATGTCAGCCTTGCCGA 2291
1016 AspGlyGluAsnSerAspSerAspAlaThrAsnAsnAlaAsnLeuThrIl 1032

2292 TCACGCT.....CATTTAAATCTCACAGGACTT. 2319
1032 eLysThrLysGluLeuLysLeuThrGlnAspLeuAsnIleSerGlyPheA 1049

2320 .....GCCACACTCAACGGC 2334
1049 snLysAlaGluIleThrAlaLysAspGlySerAspLeuThrIleGlyAsn 1065

2335 AATCTTAGTGCAGGCGGAGACACGCACTATACGGTTACGCGCAACGCCAC 2384
1066 ThrAsnSerAlaAspGlyThrAsnAlaLysLysValThrPheAsnGlnVa 1082

2385 CCAAAACGGCAACCTCAGCCTCGTGGGCAATGCCCAAGCAACATTTAATC 2434
1082 lLysAspSerLysIleSerAlaAspGlyHisLysValThrLeuHisSerL 1099

2435 AAGCCACATTAAACGGCAACACATCGGCTTCGGACAATGCTTCATTTAAT 2484
1099 ysValGluThrSerGlySerAsnAsnAsnThrGluAspSerSerAspAsn 1115

2485 CTAAGCAACAACGCCGTACAAAACGGCAGTCTGACGCTTTCCGACAACGC 2534
1116 AsnAlaGlyLeuThrIleAspAlaLysAsnValThrValAsnAsnAsnIl 1132

2535 TAAGGCAAAC......GTAAGCCATTCCGCACTCAACGGCAATGTC.... 2574

1132 eThrSerHisLysAlaValSerIleSerAlaThrSerGlyGluIleThrT 1149

2575 ..........TCCCTAGCCGATAAGGCAGTATTCCATTTTGAAAACAGC 2613
1149 hrLysThrGlyThrThrIleAsnAlaThrThrGlyAsnValGluIleThr 1165

2614 CGCTTTACCGGAAAAATCAGCGGCGGCAAGGATACG.....GCATT 2654
1166 AlaGlnThrGlySerIleLeuGlyGlyIleGluSerSerSerGlySerVa 1182

2655 ACACTTAAAAGACAGCGAATGGACGCTGCCGTCGGGCACGGAATTAGGCA 2704
1182 lThrLeuThrAlaThrGluGlyAlaLeu.....AlaValSerA 1195

2705 ATTTAAACCTTGACAACGCCACCATTACACTCAATTCCGCCTATCGACAC 2754
1195 snIleSerGlyAsnThrValThrValThrAlaAsnSer..... 1207

2755 GATGCGGCAGGCGCGCAAACCGGCAGTGCGGCAGATGCGCCGCGCCGCCG 2804
1208 .....GlyAlaLeuThrThrLeuAlaGlySer..... 1216

2805 TTCGCGCCGTTCCCTATTATCCGTTACGCCGCCAACTTCGGCAGAATCCC 2854
1217 .ThrIleLysGlyThrGluSerValThr.....ThrSerSerGlnSerG 1231

2855 GTTTCAACACGCTGACGGTAAACGGCAAATTGAACGGTCAGGGAACA... 2901
1231 ly.....AspIleGlyGlyThrIleSerGly...GlyThrVal 1242

2902 ...TTCCGCTTTATGTCGGAACTCTTCGGCTACCGCAGCGGCAAATTGAA 2948
1243 GluValLysAlaThrGluSerLeuThrThrGlnSerAsnSerLysIleLy 1259

2949 GCTGGCGGAAAGTTCCGAAGGCACTTACACCTTGGCTGTC..... 2988
1259 s...AlaThrThrGlyGluAlaAsnValThrSerAlaThrGlyThrIleG 1275

2989 .............AACAATACCGGCAACGAACCCGTAAGTCTCGAGCAA 3024
1275 lyGlyThrIleSerGlyAsnThrValAsnValThrAlaAsnAlaGlyAsp 1291

3025 TTGACGGTA.....GTGGAAGGAAAAGACAA 3050
1292 LeuThrValGlyAsnGlyAlaGluIleAsnAlaThrGluGlyAlaAlaTh 1308

3051 CACACCGCTGTCCGAAAATCTTAATTTCACCCTGCAAAACGAACACGTCG 3100
1308 rLeuThrThrSerSerGlyLysLeuThrThrGluAlaSerSerHisIleT 1325

3101 ATGCCGGCGCATGGCGTTATCAGCTTATCCGCAAAGACGGCGAGTTCCGC 3150
1325 hrSer.AlaLysGlyGlnValAsnLeuSerAlaGlnAspGlySerValAl 1341

3151 CTGCATAATCCGGTCAAAGAACAAGAGCTTTCCGACA 3187
1341 aGlySerIleAsnAlaAlaAsnValThrLeuAsnThr 1353
```

seq_name: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:US-08-617-697-2

seq_documentation_block:
; Sequence 2, Application US/08617697
; Patent No. 5977336
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen J
; TITLE OF INVENTION: High Molecular Weight Surface Proteins
; TITLE OF INVENTION: of No. 5977336-Typeable Haemophilus
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Mattare, Ltd.
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; STREET: Bldg. 1
; CITY: Arlington

```
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22202-0286
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/617,697
;      FILING DATE:  01-APR-1996
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/302,832
;      FILING DATE:  05-OCT-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US PCT/US93/02166
;      FILING DATE:  16-MAR-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Berkstresser, Jerry W
;      REGISTRATION NUMBER:  22,651
;      REFERENCE/DOCKET NUMBER:  1038-557
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 415-0810
;      TELEFAX:  (703) 415-0813
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1536 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-617-697-2


alignment_scores:
            Quality:  265.00              Length:   1297
              Ratio:  0.438                Gaps:     73
 Percent Similarity:  46.646   Percent Identity: 21.126

alignment_block:
 US-09-303-518D-653 x US-08-617-697-2   ..

 Align seg 1/1  to: US-08-617-697-2  from: 1  to: 1536

     178 AATAAAGGCAAGTTTGCAGTCGGGGCGAAAGATATTGAGGTTTACAACAA 227
         |||   |||   ::::::|||||||   ||||||   :::|||
     199 AsnHisGlyLeuIleThrValGly...LysAspGlySerVal........ 211

     228 AAAAGGGGAGTTGGTCGGCAAATCGATGACGAAAGCCCCGATGATTGATT 277
               :::|||:::|||   ::::::   :::
     212 .......AsnLeuIleGlyGlyLysValLysAsnGlu............. 221

     278 TTTCTGTGGTATCGCGTAACGGC......GTGGCGGCATTGGCGGGCGAT 321
           :::|||:::|||   ||||||      ::::::   ||||||||||:::
     222 ..GlyValIleSerValAsnGlyGlySerIleSerLeuLeuAlaGlyGln 237

     322 CAA.....................................TATATTGTGAG 335
         :::                                   |||   :::::
     238 LysIleThrIleSerAspIleIleAsnProThrIleThrTyrSerIleAl 254

     336 CGTGGCACATAACGGCGGCTATAACAATGTTGATTTTGGTGCGGAGGGAA 385
         ::::   :::|||   :::   |||      |||:::   |||:::|||:
     254 aAlaProGluAsnGluAlaValAsnLeuGlyAspIlePheAlaLysGlyG 271

     386 GCAAT.......................CCCGATCAGCACCGCTTTTCT 411
         ::|||                       :::|||   :::::::|||
     271 lyAsnIleAsnValArgAlaAlaThrIleArgAsnGlnGlyLysLeuSer 287

     412 TACCAAATTGTGAAAAGAAATAATTATAAAGCAGGGACTAACGGCCATCC 461
           :::   |||:::::::::      |||:::|||
     288 AlaAspSerValSerLysAsp......LysSerGly.............. 297

     462 TTATGGCGGCGATTATCATATGCCGCGTTTGCACAAATTTGTCACAGATG 511
                                           :::::::|||   :::|
     298 ..............................AsnIleValLeuSerA 303

     512 CAGAACCTGTTGAG.................ATGACCAGTTATATGGAT 543
         ||:::      |||                 :::::::::      :::
     303 laLysGluGlyGluAlaGluIleGlyGlyValIleSerAlaGlnAsnGln 319

     544 GGGTGGAAATACGCTGATTTAAATAAATACCCTGATCGTGTT......CG 587
               |||   :::   |||           |||:::|||      ::
     320 GlnAlaLysGlyGlyLysLeuMetIleThrGlyAspLysValThrLeuLy 336

     588 AATCGGAGCA.........................GGCAGACAATATT 610
         :   ||||||                        |||:::   |||
     336 sThrGlyAlaValIleAspLeuSerGlyLysGluGlyGlyGluThrTyrL 353

     611 GGCGGTCTGATGAA...GACGAACCCAATAAC.................. 639
               :::||||||      |||   :::|||
     353 euGlyGlyAspGluArgGlyGluGlyLysAsnGlyIleGlnLeuAlaLys 369

     640 ..............CGCGAAAGTTCATATCATATTGCAAGC........ 666
                      :::   |||:::   ::::::::::::
     370 LysThrSerLeuGluLysGlySerThrIleAsnValSerGlyLysGluLy 386

     667 ..........GCATATTCTTGG............CTCGTCGGTGGCAATA 694
                    |||      |||           |||:::   ||||||
     386 sGlyGlyArgAlaIleValTrpGlyAspIleAlaLeuIleAspGlyAsnI 403

     695 CCTTTGCACAAAATGGATCAGGTGGTGGCACAGTCAACTTAGGTAGCGAA 744
               ||||||   |||||||||
     403 leAsnAlaGln...GlySerGly........................... 409

     745 AAAATTAAACATAGCCCATATGGTTTTTTACCAACAGGAGGCTCATTTGG 794
                                      :::      |||||||||
     410 ....................AspIleAlaLysThrGlyGlyPheValGl 419

     795 CGACAGTGGCTCACCAATGTTTATCTATGATGCCCAAAAGCAAAAGTGGT 844
         ||||||      :::||||||   |||
     419 uThrSerGlyHisAspLeuPheIleLysAsp................... 429

     845 TAATTAATGGGGTATTGCAAACAGGCAACCCCTATATAGGAAAAAGCAAT 894
               |||:::::::::::::::
     430 .....AsnAlaIleValAspAla........................... 435

     895 GGCTTCCAGCTAGTTCGTAAAGATTGGTTCTATGATGAAATCTTTGCTGG 944
                           |||:::|||:::   |||
     436 ..................LysGluTrpLeuLeuAsp............Ph 442

     945 AGATACCCATTCAGTATTCTACGAACCACATCAAAATGGGAAATACTTTT 994
          |||:::   |||:::      |||         :::::::::
     442 eAspAsnValSerIleAsnAlaGluThrAlaGlyArgSerAsnThrSerG 459

     995 TTAACGACAATAATAATGGCGCAGGAAAAATCGATGCCAAACATAAACAC 1044
           :::|||:::   :::|||:::|||                    :::
     459 luAspAspGluTyrThrGlySerGly..................AsnSer 469

    1045 TATTCTCTACCTTATAGATTAAAAACACGAACCGTTCAATTGTTTAATGT 1094
            |||   |||   |||   |||   :::|||:::   |||   |||::
     470 AlaSerThrProLysArgAsnLysGluLysThrThr...LeuThrAsnTh 485

    1095 TTCTTTATCCGAG.............................ACAGCAA 1114
         ::::|||::::::                             |||||||:
     485 rThrLeuGluSerIleLeuLysLysGlyThrPheValAsnIleThrAlaA 502

    1115 GAGAACCTGTTTAT................................. 1128
         :::::   :::|||
     502 snGlnArgIleTyrValAsnSerSerIleAsnLeuSerAsnGlySerLeu 518

    1129 ..................CATGCTGCAGGTGGGGTCAACAGTTATCGACC 1160
                           :::::::::|||||||||
     519 ThrLeuTrpSerGluGlyArgSerGlyGlyGlyVal.............. 530

    1161 CAGACTGAATAAT.........................GGAGAAAATA 1183
```

```
          ::::::||||||                                   |||    |||:
 531  .GluIleAsnAsnAspIleThrThrGlyAspAspThrArgGlyAlaAsnL 547

1184  TTTCCTTTATTGACAAAGGAAAAGGTGAATTGATACTTACCAGCAACATC 1233
      :::::::   :::   |||       :::      :::   :::||||||
 547  euThrIleTyrSerGlyGlyTrpValAsp......ValHisLysAsnIle 561

1234  AACCAAGGCGCGGGCGGTTTGTATTTTGAGGGTAATTTTACGGTCTCGCC 1283
      :::   ||||||                  :::||||||:::::::::::
 562  SerLeuGlyAla..............GlnGlyAsnIleAsnIleThrAl 573

1284  TAAA.......................AACAACGAAACGTGGCAAGGCG 1309
       |||                        :::|||:::::      |||
 573  aLysGlnAspIleAlaPheGluLysGlySerAsnGlnValIleThrGlyG 590

1310  CGGGCGTTCATATCAGTGATGGCAGTACCGTTACTTGGAAAGTAAACGGC 1359
         |||:::   |||::::::|||:::          ::::::   |||:::
 590  lnGlyThr...IleThrSerGlyAsnGlnLysGlyPheArgPheAsnAsn 605

1360  GTGGCAAACGACCGCCTGTCCAAAATCGGCAAAGGCACGCTGCTGGTTCA 1409
      |||:::          |||:::      |||:::|||   |||
 606  ValSer.........LeuAsnGlyThrGlySerGly...LeuGlnPheTh 618

1410  AGCCAAAGGGGAAAAC.......................CAAGGCTCGG 1435
       :::|||      |||                        :::|||::::
 618  rThrLysArgThrAsnLysTyrAlaIleThrAsnLysPheGluGlyThrL 635

1436  TCAGCGTGGGCGACGGTAAAGTCATCTTA.................GAT 1467
      :::::::   :::|||||||||   :::                   :::
 635  euAsnIle...SerGlyLysValAsnIleSerMetValLeuProLysAsn 650

1468  CAGCAGGCGGACGATCAAGGCAAAAAACAAGCCTTT............. 1503
      :::::::::   |||:::   |||   :::::::::
 651  GluSerGlyTyrAspLysPheLysGlyArgThrTyrTrpAsnLeuThrSe 667

1504  ..........AGTGAAATCGGC...........TTGGTCAGCGGCAGGG 1531
                |||||    |||            :::::::::||||
 667  rLeuAsnValSerGluSerGlyGluPheAsnLeuThrIleAspSerArgG 684

1532  GGACG.......................................GTG 1539
      ||:::                                        :::
 684  lySerAspSerAlaGlyThrLeuThrGlnProTyrAsnLeuAsnGlyIle 700

1540  CAACTGAATGCCGATAATCAGTTCAACCCCGACAAA............. 1575
      :::::::|||   |||:::   ||||||   ::::::
 701  SerPheAsnLysAspThrThrPheAsnValGluArgAsnAlaArgValAs 717

1576  .............................................CTCT 1579
                                                   |
 717  nPheAspIleLysAlaProIleGlyIleAsnLysTyrSerSerLeuAsnT 734

1580  ATTTCGGCTTTCGCGGCGGACGTTTGGATTTGAACGGGCATTCGCTTTCG 1629
       ||   :::|||:::|||:::             :::|||   |||::::::
 734  yrAlaSerPheAsnGlyAsnIleSerValSerGlyGlyGlySerValAsp 750

1630  TTCCACCGCATT.........CAAAATACCGATGAAGGGGCGATGATTGT 1670
       |||      :::         :::|||::::::     :::::::::::
 751  PheThrLeuLeuAlaSerSerSerAsnValGlnThrProGlyValValIl 767

1671  CAACCACAAT.............................CAAGACAAAG 1690
      :|||     :::                             :::    |||
 767  eAsnSerLysTyrPheAsnValSerThrGlySerSerLeuArgPheLysT 784

1691  AATCCACCGTTACCATTACAGGC.........AATAAAGATATTACT... 1728
        |||      |||     ||||||          :::||||||:::|||
 784  hrSerGlySerThrLysThrGlyPheSerIleGluLysAspLeuThrLeu 800

1729  ...ACAACCGGCAATAAC............................... 1743
          :::|||||||:::|||
 801  AsnAlaThrGlyGlyAsnIleThrLeuLeuGlnValGluGlyThrAspGl 817

1744  ..........AACAACTTGGATAGCAAAAAAGAAATTGCCTACAACGGT. 1782
                 :::::::::   :::||||||:::|||:::::::::|||

 817  yMetIleGlyLysGlyIleValAlaLysLysAsnIleThrPheGluGlyG 834

1783  ........TGGTTTGGCGAGAAAGATGCA...ACCAAAACGAACGGGCGG 1821
                 ||||||::::::   |||   |||:::   :::|||:::
 834  lyAsnIleThrPheGlySerArgLysAlaValThrGluIleGluGlyAsn 850

1822  CTCAATCTGAATTACCAACCGGAAGAAGCGGATCGCACTTTA........ 1863
      :::::::::|||             :::::::|||:::   ||||||
 851  ValThrIleAsn.........AsnAsnAlaAsnValThrLeuIleGlySe 864

1863  .................................................. 1863

 864  rAspPheAspAsnHisGlnLysProLeuThrIleLysLysAspValIleI 881

1864  ..............CTGCTTTCCGGCGGA......ACAAATTTAAACGGC 1893
                     |||   :::||||||      :::|||:::   |||
 881  leAsnSerGlyAsnLeuThrAlaGlyGlyAsnIleValAsnIleAlaGly 897

1894  AATATCACG...CAAACAAACGGCAAACTGTTTTTCAGCGGCA....... 1933
      |||:::|||   ::::::|||  ||||||   |||:::|||
 898  AsnLeuThrValGluSerAsn.AlaAsn...PheLysAlaIleThrAsnP 913

1934  ..........................GACCGACACCGCACGCCTACAATC 1957
                               |||:::       :::    :::|||
 913  heThrPheAsnValGlyGlyLeuPheAspAsnLysGlyAsnSerAsnIle 929

1958  ATTTAGGAAGCGGGTGGTCAAAAATGGAAGGTATCCCACAAGGAGAAATC 2007
                  ||||||   ::::::|||           ::::::|||::
 930  SerIleAlaLysGlyGlyAlaArgPheLysAspIleAspAsnSerLysAs 946

2008  GTGTGGGACAACGATTGGATCGACCGC.ACATTTAAAGCGGAAAACTTCC 2056
        :    :::    |||   :::|||:::   |||::::::
 946  nLeuSerIleThrThrAsnSerSerSerThrTyrArg............ 958

2057  ATATTCAGGGCGGACAAGCGGTGGTTTCCCGCAATGTTGCCAAAGTGGAA 2106
                        :::::::::|||    |||:::::::::    :::
 959  ..................ThrIleIleSerGlyAsnIleThrAsnLysAsn 969

2107  GGCGATTGGCATTTAAGCAATCACGCC............CAAGCAGTTTT 2144
      ||||||    :::::::::|||::::::               :::    :::
 970  GlyAspLeuAsnIleThrAsnGluGlySerAspThrGluMetGlnIleGl 986

2145  CGGTGTCGCACCGCATCAAAGCCACACAATCTGTACACGTTCGGACTGGA 2194
        |||    :::    :::::::::    ::::::           ||||||
 986  yGlyAspValSerGlnLysGluGlyAsnLeuThrIleSerSerAspLysI 1003

2195  CGGGTCTGACAAGTTGTACCGAAAAAACCATTACC...GACGATAAAGTG 2241
        ::::::|||            |||    ||||||            |||
1003  leAsnIleThr...........LysGlnIleThrIleLysAlaGlyVal 1015

2242  ATTGCTTCATTGAGCAAGACCGACATCAGAGGCAATGTCAGCCTTGCCGA 2291
         ::::::   |||   :::|||      :::|||:::::|||:::
1016  AspGlyGluAsnSerAspSerAspAlaThrAsnAsnAlaAsnLeuThrIl 1032

2292  TCACGCT....................CATTTAAATCTCACAGGACTT. 2319
        :::                            ||||||:::::|||:::
1032  eLysThrLysGluLeuLysLeuThrGlnAspLeuAsnIleSerGlyPheA 1049

2320  ................................GCCACACTCAACGGC 2334
                                        |||:::::::::
1049  snLysAlaGluIleThrAlaLysAspGlySerAspLeuThrIleGlyAsn 1065

2335  AATCTTAGTGCAGGCGGAGACACGCACTATACGGTTACGCGCAACGCCAC 2384
      :::    ||||||   |||    :::             ||||||   |||   ::
1066  ThrAsnSerAlaAspGlyThrAsnAlaLysLysValThrPheAsnGlnVa 1082

2385  CCAAAACGGCAACCTCAGCCTCGTGGGCAATGCCCAAGCAACATTTAATC 2434
      ::::::::::::::|||       |||:::      :::      ::::
1082  lLysAspSerLysIleSerAlaAspGlyHisLysValThrLeuHisSerL 1099

2435  AAGCCACATTAAACGGCAACACATCGGCTTCGGACAATGCTTCATTTAAT 2484
      :::::       :::|||:::::::::   ::::::::::::|||    |||
1099  ysValGluThrSerGlySerAsnAsnAsnThrGluAspSerSerAspAsn 1115
```

```
 2485 CTAAGCAACAACGCCGTACAAAACGGCAGTCTGACGCTTTCCGACAACGC 2534
      ::::::  :::::::::      ::::::|||:::::::::|||
 1116 AsnAlaGlyLeuThrIleAspAlaLysAsnValThrValAsnAsnAsnIl 1132

 2535 TAAGGCAAAC......GTAAGCCATTCCGCACTCAACGGCAATGTC.... 2574
      ::::::      ||||||  ||||||  :::|||:::::
 1132 eThrSerHisLysAlaValSerIleSerAlaThrSerGlyGluIleThrT 1149

 2575 ..........TCCCTAGCCGATAAGGCAGTATTCCATTTTGAAAACAGC 2613
            :::   :::  ::::::  :::  |||  :::
 1149 hrLysThrGlyThrThrIleAsnAlaThrThrGlyAsnValGluIleThr 1165

 2614 CGCTTTACCGGAAAAATCAGCGGCGGCAAGGATACG.........GCATT 2654
         ||||||:::|||  ||||||  ::::::      :::::
 1166 AlaGlnThrGlySerIleLeuGlyGlyIleGluSerSerSerGlySerVa 1182

 2655 ACACTTAAAAGACAGCGAATGGACGCTGCCGTCGGGCACGGAATTAGGCA 2704
      :  |||      :::|||  :::|||              ::::::|
 1182 lThrLeuThrAlaThrGluGlyAlaLeu............AlaValSerA 1195

 2705 ATTTAAACCTTGACAACGCCACCATTACACTCAATTCCGCCTATCGACAC 2754
      ||:::::::  :::::::::|||:::|||  ||||||
 1195 snIleSerGlyAsnThrValThrValThrAlaAsnSer............ 1207

 2755 GATGCGGCAGGCGCGCAAACCGGCAGTGCGGCAGATGCGCCGCGCCGCCG 2804
             ||||||  |||      |||::::::
 1208 .........GlyAlaLeuThrThrLeuAlaGlySer.............. 1216

 2805 TTCGCGCCGTTCCCTATTATCCGTTACGCCGCCAACTTCGGCAGAATCCC 2854
       :::  ::::::      |||||||||      ||||||::::::|||
 1217 .ThrIleLysGlyThrGluSerValThr......ThrSerSerGlnSerG 1231

 2855 GTTTCAACACGCTGACGGTAAACGGCAAATTGAACGGTCAGGGAACA... 2901
                       ::::::|||  :::::::|||  ||||||
 1231 ly............AspIleGlyGlyThrIleSerGly...GlyThrVal 1242

 2902 ...TTCCGCTTTATGTCGGAACTCTTCGGCTACCGCAGCGGCAAATTGAA 2948
         :::      ::::::|||            ::::::|||:::||
 1243 GluValLysAlaThrGluSerLeuThrThrGlnSerAsnSerLysIleLy 1259

 2949 GCTGGCGGAAAGTTCCGAAGGCACTTACACCTTGGCTGTC.......... 2988
      |  |||  ::::::|||::::::  |||  |||:::
 1259 s...AlaThrThrGlyGluAlaAsnValThrSerAlaThrGlyThrIleG 1275

 2989 ..............AACAATACCGGCAACGAACCCGTAAGTCTCGAGCAA 3024
                     :::||||||  |||      ::::::      :::
 1275 lyGlyThrIleSerGlyAsnThrValAsnValThrAlaAsnAlaGlyAsp 1291

 3025 TTGACGGTA.......................GTGGAAGGAAAAGACAA 3050
      |||||||||                      :::||||||      ::
 1292 LeuThrValGlyAsnGlyAlaGluIleAsnAlaThrGluGlyAlaAlaTh 1308

 3051 CACACCGCTGTCCGAAAATCTTAATTTCACCCTGCAAAACGAACACGTCG 3100
      :       |||::::::        |||      ::::::|||:::
 1308 rLeuThrThrSerSerGlyLysLeuThrThrGluAlaSerSerHisIleT 1325

 3101 ATGCCGGCGCATGGCGTTATCAGCTTATCCGCAAAGACGGCGAGTTCCGC 3150
        ::: |||   |||  ::::::|||||||||:::   :::|||   ||
 1325 hrSer.AlaLysGlyGlnValAsnLeuSerAlaGlnAspGlySerValAl 1341

 3151 CTGCATAATCCGGTCAAAGAACAAGAGCTTTCCGACA 3187
      |      |||::::::   |||   ::::::   |||
 1341 aGlySerIleAsnAlaAlaAsnValThrLeuAsnThr 1353

seq_name: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-08-719-641-2

seq_documentation_block:
; Sequence 2, Application US/08719641
; Patent No. 6218141
;  GENERAL INFORMATION:
;    APPLICANT:  Barenkamp, Stephen J
;    TITLE OF INVENTION:  High Molecular Weight Surface Proteins
;    TITLE OF INVENTION:  of No. 6218141-Typeable Haemophilus
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Shoemaker and Mattare, Ltd.
;      STREET:  2001 Jefferson Davis Hwy., 1203 Crystal Plaza
;      STREET:  Bldg. 1
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22202-0286
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/719,641
;      FILING DATE:
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/302,832
;      FILING DATE:  16-SEP-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US PCT/US93/02166
;      FILING DATE:  16-MAR-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9205704.1
;      FILING DATE:  16-MAR-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Berkstresser, Jerry W
;      REGISTRATION NUMBER:  22,651
;      REFERENCE/DOCKET NUMBER:  1038-625
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 415-0810
;      TELEFAX:  (703) 415-0813
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1536 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:
US-08-719-641-2


alignment_scores:
            Quality:  265.00              Length:   1297
              Ratio:  0.438                 Gaps:     73
 Percent Similarity:  46.646   Percent Identity: 21.126

alignment_block:
 US-09-303-518D-653 x US-08-719-641-2   ..

 Align seq 1/1  to: US-08-719-641-2  from: 1  to: 1536

     178 AATAAAGGCAAGTTTGCAGTCGGGGCGAAAGATATTGAGGTTTACAACAA 227
         |||   |||   :::::||||||   ||||||   :::|||
     199 AsnHisGlyLeuIleThrValGly...LysAspGlySerVal........ 211

     228 AAAAGGGGAGTTGGTCGGCAAATCGATGACGAAAGCCCCGATGATTGATT 277
                ::::|||:::|||   ::::::   :::
     212 .......AsnLeuIleGlyGlyLysValLysAsnGlu.............. 221

     278 TTTCTGTGGTATCGCGTAACGGC......GTGGCGGCATTGGCGGGCGAT 321
           :::|||:::|||   ||||||      ::::::   |||||||||:::
     222 ..GlyValIleSerValAsnGlyGlySerIleSerLeuLeuAlaGlyGln 237

     322 CAA.....................................TATATTGTGAG 335
         :::                                    |||   :::::
     238 LysIleThrIleSerAspIleIleAsnProThrIleThrTyrSerIleAl 254

     336 CGTGGCACATAACGGCGGCTATAACAATGTTGATTTTGGTGCGGAGGGAA 385
         ::::   :::|||   :::   |||      |||:::   |||:::|||:
```

```
    1341 aGlySerIleAsnAlaAlaAsnValThrLeuAsnThr 1353

seq_name: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-09-268-347-36

seq_documentation_block:
; Sequence 36, Application US/09268347
; Patent No. 6335182
; GENERAL INFORMATION:
;  APPLICANT: Loosmore, Sheena M.
;  TITLE OF INVENTION: RECOMBINANT HAEMOPHILUS INFLUENZAE ADHESIN PROTEINS
;  FILE REFERENCE: 1038-860
;  CURRENT APPLICATION NUMBER: US/09/268,347
;  CURRENT FILING DATE: 1999-03-16
;  NUMBER OF SEQ ID NOS: 54
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 36
;   LENGTH: 2411
;   TYPE: PRT
;   ORGANISM: Haemophilus influenzae
US-09-268-347-36


alignment_scores:
            Quality:  257.50              Length:   1701
              Ratio:  0.341                Gaps:      75
 Percent Similarity:  44.444   Percent Identity: 18.871

alignment_block:
 US-09-303-518D-653 x US-09-268-347-36   ..

 Align seg 1/1  to: US-09-268-347-36  from: 1  to: 2411

       7 ACAACCGACAAACGGACAACCGAA.........ACACACCGCAAAGCCCC 47
         |||   ||||||:::|||   :::          |||   ::::::
     242 ThrGlyAspLysAsnThrLeuAspValValLeuThrAlaLysGluAsnGl 258

      48 TAAAACCGGCCGCATCCGCTTCTCGCCCGCTTACTTAGCCATATGCCTGT 97
         ||||||   ::::::::::|||:::|||          :::|||      :
     258 yLysThrThrGluValLysPheThrProLysThrSerValIleLysGluL 275

      98 CGTTCGGCATTCTGCCCCAAGCCCGGGCGGGACACACTTATTTCGGCATC 147
         ::   |||   |||      ::::::   ::::::                :::
     275 ysAspGlyLysLeuPheThrGlyLysGluAsnAsnAspThrAsnLysVal 291

     148 AACTACCAATACTATCGCGACTTTGCCGAAAATAAAGGCAAGTTTGCAGT 197
         :::   :::          |||   :::::::::   :::   ::::::::
     292 ThrSerAsnThrAlaThrAspAsnThrAspGluGlyAsnGlyLeuValTh 308

     198 CGGGGCGAAAGATATTGAGGTTTACAACAAAAAAGGGGAGTTGGTCGGCA 247
         ::::         |||::::::   ||||||   |||      |||
     308 rAlaLysAlaValIleAspAlaValAsnLysAlaGlyTrpArgValLysT 325

     248 AATCGATGACGAAAGCCCCGATGATTGATTTT...TCTGTGGTATCGCGT 294
           :::   :::::::::         ||||||   :::|||:::|||
     325 hrThrThrAlaAsnGlyGlnAsnGlyAspPheAlaThrValAlaSerGly 341

     295 AACGGCGTGGCGGCATTGGCGGGCGATCAATATATTGTGAGCGTGGCACA 344
         ::::::|||:::      :::||||||         :::||||||:::
     342 ThrAsnValThrPheGluSerGlyAspGlyThrThrAlaSerValThrLy 358

     345 TAACGGCGGCTATAACAATGTTGATTTTGGTGCGGAGGGAAGCAATCCCG 394
         :::   :::   |||:::
     358 sAspThrAsnGlyAsnGly............................... 364

     395 ATCAGCACCGCTTTTCTTACCAAATTGTGAAAAGAAATAATTATAAAGCA 444
                               |||:::   :::   :::   |||:::
     365 ......................IleThrValLysTyrAspAlaLysVal 373

     445 GGGACTAACGGCCATCCTTATGGCGGCGATTATCATATGCCGCGTTTGCA 494
         |||   :::|||      :::   :::|||
     374 Gly...AspGlyLeuLysPheAspSerAsp.................Ly 383

     495 CAAATTTGTCACAGATGCAGAACCTGTTGAGATGACCAGTTATATGGATG 544
```

```
         |||:::|||:::|||:::      :::   :::|||           |
     383 sLysIleValAlaAspThrThrAlaLeuThrValThr.....G 396

     545 GGTGGAAATACGCTGATTTAAATAAATACCCTGATCGTGTTCGAATC... 591
         ||   |||   |||::::::   |||      |||:::   ::::::
     396 lyGlyLysValAlaGluIleAlaLysGluAspAspLysLysLysLeuVal 412

     592 .....GGAGCAGGCAGACAATATTGGCGGTCTGA 620
                                  :::   |||:::      |||::::::
     413 AsnAlaGlyAspLeuValThrAlaLeuGlyAsnLeuSerTrpLysAlaLy 429

     621 TGAAGACGAACCCAATAACCGCGAAAGTTCATATCATATTGCAAGCGCAT 670
             :::      ::::::   :::::::::   :::          |||
     429 sAlaGluAlaAspThrAspThrAspGlyAlaLeuGluGlyIleSerLysA 446

     671 ATTCTTGGCTCGTCGGTGGCAATACCTTTGCA..................... 702
           :::   :::   :::|||:::|||   :::
     446 spGlnGluValLysAlaGlyGluThrValThrPheLysAlaGlyLysAsn 462

     703 ............CAAAATGGATCAGGT..... 717
                     |||:::|||::::::
     463 LeuLysValLysGlnAspGlyAlaAsnPheThrTyrSerLeuGlnAspAl 479

     718 ............................GGTGGCACAGTCAACTTAGGTA 739
                                     |||||||||:::|||   |||:
     479 aLeuThrGlyLeuThrSerIleThrLeuGlyGlyThrThrAsnGlyGlyA 496

     740 GCGAAAAA.........ATTAAACATAGCCCATATGGTTTTTTACCAACA 780
         :::::             |||:::   :::          :::   |||:::
     496 snAspAlaLysThrValIleAsnLysAspGlyLeuThrIleThrProAla 512

     781 GGAGGCTCATTTGGCGACAGTGGCTCACCAATGTTTATCTATGATGCCCA 830
         |||::::::   |||   :::|||:::      :::
     513 GlyAsnGly...GlyThrThrGlyThrAsnThrIleSer..... 524

     831 AAAGCAAAAGTGGTTAATTAATGGGGTATTGCAAACAGGCAACCCCTATA 880
                     :::   :::      :::::::::||||||      |
     525 .............ValThrLysAspGlyIleLysAlaGlyAsnLysAlaI 537

     881 TA...GGAAAAAGCAATGGCTTCCAGCTAGTTCGTAAAGATTGGTTCTAT 927
         ||   :::   ::::::|||::::::                  |||
     537 leThrAsnValAlaSerGlyLeuArgAla.................Tyr 547

     928 GATGAAATCTTTGCTGGAGATACCCATTCAGTATTCTACGAACCACATCA 977
         |||:::
     548 AspAspAla.......................................... 550

     978 AAATGGGAAATACTTTTTTAACGACAATAATAATGGCGCAGGAAAAATCG 1027
                       |||:::      ||||||:::|||      :::
     551 .............AsnPheAspValLeuAsnAsnSerAlaThrAspLeu. 562

    1028 ATGCCAAACATAAACACTATTCTCTACCTTATAGATTAAAAACACGAACC 1077
                 :::::::|||   :::      |||:::
     563 ........AsnArgHisValGluAspAlaTyrLys.............. 571

    1078 GTTCAATTGTTTAATGTTTCTTTATCCGAGACAGCAAGAGAACCT..... 1122
               |||:::|||::::::   :::::::::   :::::::|||
     572 ...GlyLeuLeuAsnLeuAsnGluLysAsnAlaAsnLysGlnProLeuVa 587

    1123 ..............................................G 1123
                                                      |
     587 lThrAspSerThrAlaAlaThrValGlyAspLeuArgLysLeuGlyTrpV 604

    1124 TTTATCATGCTGCAGGTGGGGTCAACAGTTATCGACCCAGACTGAATAAT 1173
         ||      :::   :::|||:::::::::         ::::::::::::
     604 alValSerThrLysAsnGlyThrLysGluGluSerAsnGlnValLysGln 620

    1174 GGAGAAAATATTTCCTTTATTGACAAAGGAAAAGGTGAATTGATACTTAC 1223
         ::::::::::::   |||          |||   |||          :::||
     621 AlaAspGluValLeuPhe......ThrGlyAlaGlyAlaAlaThrValTh 635

    1224 CAGCAACATCAACCAAGGCGCGGGCGGTTTGTATTTTGAGGGTAATTTTA 1273
         ||||:::   :::                          :::|||:::   |
```

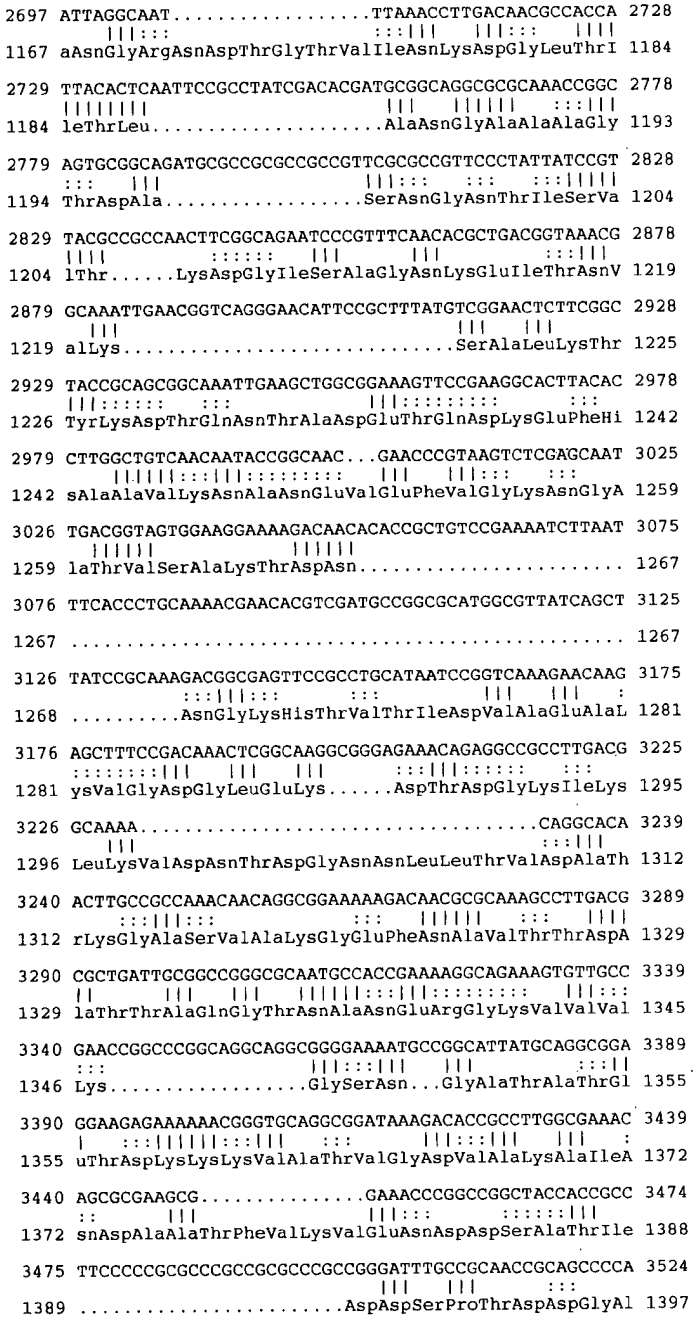

635 rSerLysSerGlu.....AsnGlyLysHisT 644
1274 CGGTCTCGCCTAAAAACAACGAAACGTGGCAAGGCGCGGGCGTTCATATC 1323
644 hrIleThrValSerValAlaGluThrLysAlaAspSerGlyLeuGlu... 659
1324 AGTGATGGCAGTACCGTTACTTGGAAAGTAAACGGCGTGGCAAACGACCG 1373
660 LysAspGlyAspThrIleLysLeuLysValAspAsnGlnAsnThrAspAs 676
1374 CCTGTCCAAAATCGGC...AAAGGCACGCTGCTGGTTCAAGCCAAAGGGG 1420
676 nValLeuThrValGlyAsnAsnGlyThrAlaValThrLysGlyGlyPheG 693
1421 AA......AACCAAGGCTCGGTCAGCGTGGGCGACGGTAAAGTC...ATC 1461
693 luThrValLysThrGlyAlaThrAspAlaAspArgGlyLysValThrVal 709
1462 TTAGATCAGCAGGCGGACGATCAAGGCAAAAAACAAGCC...TTTAGTGA 1508
710 LysAspAlaThrAlaAsnAspAlaAspLysLysValAlaThrValLysAs 726
1509 AATCGGCTTGGTCAGCGGCAGGGGGACGGTGCAACTGAATGCCGATAAT. 1557
726 pValAlaThrAlaIleAsnSerAlaAlaThrPheValLysThrGluAsnL 743
1558 .....CAGTTCAACCCC.....GAC 1572
743 euThrThrSerIleAspGluAspAsnProThrAspAsnGlyLysAspAsp 759
1573 AAACTCTATTTCGGC.....TTTCGCGGCGGA.....CG 1601
760 AlaLeuLysAlaGlyAspThrLeuThrPheLysAlaGlyLysAsnLeuLy 776
1602 TTTGGATTTGAACGGGCATTCGCTTTCGTTCCACCGCATTCAAAATACCG 1651
776 sValLysArgAspGlyLysAsnIleThrPheAspLeuAlaLysAsnLeuG 793
1652 ATGAAGGGGCGATGATTGTCAACCACAATCAAGACAAAGAATCCACCGTT 1701
793 luValLysThrAlaLysValSer.....AspThrLeu 803
1702 ACCATTACAGGCAATAAA..... 1719
804 ThrIleGlyGlyAsnThrProThrGlyGlyThrThrAlaThrProLysVa 820
1720 .GATATTACTACAACCGGCAATAACAACAACTTG..... 1752
820 lAsnIleThrSerThrAlaAspGlyLeuAsnPheAlaLysGluThrAlaA 837
1753 .....GATAGCAAAAAAGAAATTGCCTACAACGGT.....TGGTTT 1788
837 spAlaSerGlySerLysAsnValTyrLeuLysGlyIleAlaThrThrLeu 853
1789 GGCGAGAAAGATGCAACCAAAACGAACGGGCGGCTCAATCTGAATTACCA 1838
854 ThrGluProSerAlaGlyAlaLysSerSerHisValAspLeuAsnValAs 870
1839 ACCGGAAGAAGCGGATCGCACTTTA.....CTGCTTTCCGGCG 1876
870 pAlaThrLysLysSerAsnAlaAlaSerIleGluAspValLeuArgAlaG 887
1877 GAACAAATTTAAACGGCAAT..... 1896
887 lyTrpAsnIleGlnGlyAsnGlyAsnAsnValAspTyrValAlaThrTyr 903
1897 .....AT 1898
904 AspThrValAsnPheThrAspAspSerThrGlyThrThrThrValThrVa 920
1899 CACGCAAACA..... 1908
920 lThrGlnLysAlaAspGlyLysGlyAlaAspValLysIleGlyAlaLysT 937
1909 .....AACGGCAAACTGTTTTTCAGCGGCAGACCG 1938
937 hrSerValIleLysAspHisAsnGlyLysLeu...PheThrGlyLysAsp 952
1939 ACACCGCACGCCTACAATCATTTAGGAAGCGGGTGGTCAAAAATGGAAGG 1988
953 LeuLysAspAlaAsnAsn.....GlyAlaThrValSerGluAspAspGl 967
1989 TATCCCACAAGGAGAA...ATCGTGTGGGACAACGATTGGATCGACCGCA 2035
967 yLysAspThrGlyThrGlyLeuValThrAlaLysThrValIleAspAlaV 984
2036 CATTTAAAGCGGAAAACTTCCATATTCAGGGCGGACAAGCGGTGGTTTCC 2085
984 alAsnLysSer...GlyTrpArgValThrGlyGluGlyAlaThrAlaGlu 999
2086 CGCAATGTTGCCAAAGTGGAAGGCGATTGGCATTTAAGCAATCACGCCCA 2135
1000 ThrGlyAlaThrAlaVal.....AsnAlaGlyAsnAlaGluTh 1012
2136 AGCAGTTTTCGGTGTCGCA.....CCGCATCAAAGCCACACAATCT 2176
1012 rValThrSerGlyThrSerValAsnPheLysAsnGlyAsnAlaThrThrA 1029
2177 GTACACGTTCGGACTGGACGGGT..... 2199
1029 laThrValSerLysAspAsnGlyAsnIleAsnValLysTyrAspValAsn 1045
2200 .....CTGACAAGTTG 2210
1046 ValGlyAspGlyLeuLysIleGlyAspAspLysLysIleValAlaAspTh 1062
2211 TACCGAAAAAACCATTACCGACGATAAAGTGATTGCTTCATTGAGCAAGA 2260
1062 rThrThrLeuThrValThrGlyGlyLysValSerValProAlaGlyAlaA 1079
2261 CCGACATCAGAGGCAATGTCAGCCTTGCCGATCACGCTCATTTAAATCTC 2310
1079 snSerValAsnAsnAsnLysLysLeu.....ValAsnAla 1090
2311 ACAGGACTTGCCACACTCAACGGCAATCTTAGT..... 2343
1091 GluGlyLeuAlaThrAlaLeuAsnAsnLeuSerTrpThrAlaLysAlaAs 1107
2344 .....GCAGGCGGAGACACGCACTATACGGTT.... 2370
1107 pLysTyrAlaAspGlyGluSerGluGlyGluThrAspGlnGluValLysA 1124
2371 .....ACGCGCAACGCCACCCAAAACGGCAACCTCAGCCTCGTGGGCAAT 2415
1124 laGlyAspLysValThrPheLysAlaGlyLysAsnLeuLysValLysGln 1140
2416 GCCCAAGCAACATTTAAT.....CAAGCCACATTAAACGGCAACAC 2456
1141 SerGluLysAspPheThrTyrSerLeuGlnAspThrLeuThrGlyLeuTh 1157
2457 ATCGGCTTCGGACAATGCTTCATTTAATCTAAGCAACAAC.....G 2497
1157 rSerIleThrLeuGlyGlyThrAlaAsnGlyArgAsnAspThrGlyThrV 1174
2498 CCGTACAAAACGGCAGTCTGACGCTTTCCGACAACGCTAAGGCAAACGTA 2547
1174 alIleAsnLysAspGlyLeuThrIleThrLeuAlaAsnGlyAlaAlaAla 1190
2548 AGCCATTCCGCACTCAACGGCAATGTCTCCCTAGCCGATAAGGCAGTATT 2597
1191 GlyThrAspAlaSerAsnGlyAsnThr..... 1199
2598 CCATTTTGAAAACAGCCGCTTTACCGGAAAAATCAGCGGCGGCAAGGATA 2647
1200 .....IleSerValThrLysAsp. 1205

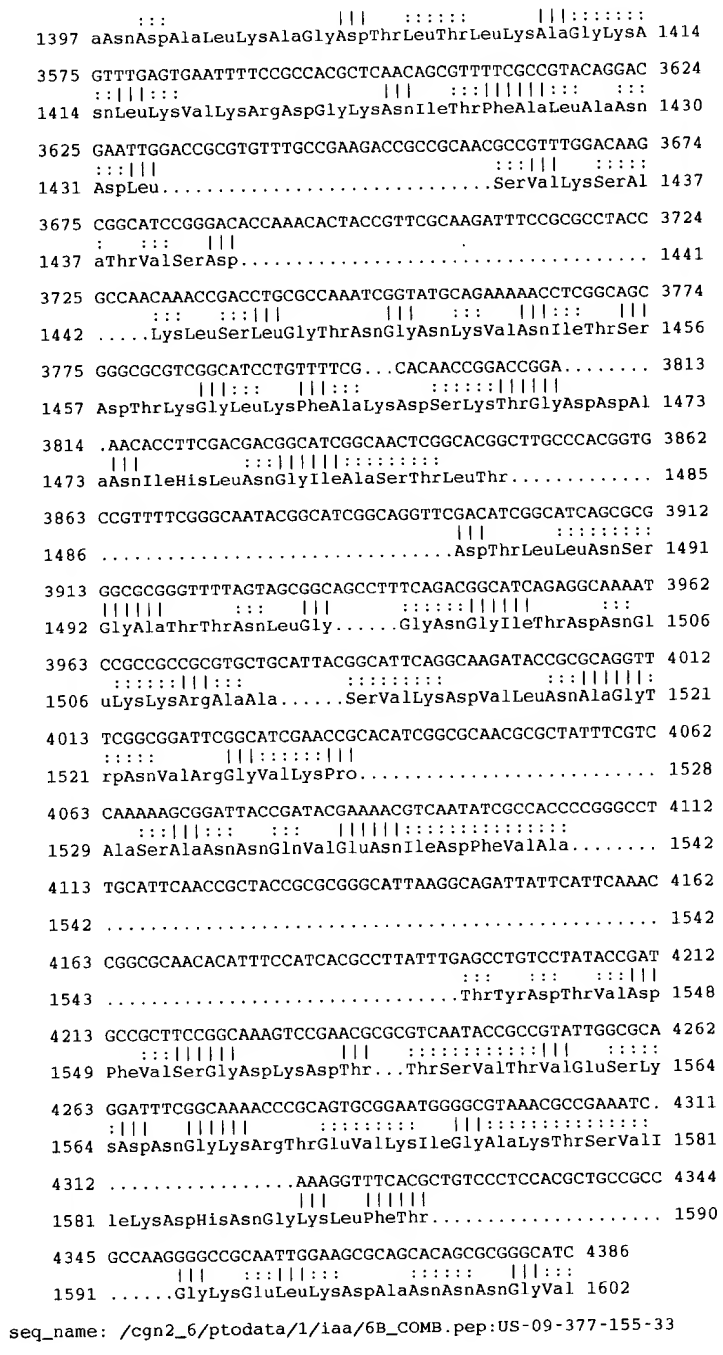
```
2648 CGGCATTACACTTAAAAGACAGCGAATGGACGCTGCCGTCGGGCACGGAA 2697
                                                |||  :::
1206 ..........................................GlyIleSer 1208

2698 TTAGGCAATTTAAACCTTGACAACGCCACCATTACACTCAATTCCGCCTA 2747
       ||||||  ::::::  |||:::      :::|||:::::::   ||
1209 AlaGlyAsnLysGluIleThrAsnValLysSerAlaLeuLysThr...Ty 1224

2748 TCGA......CACGATGCGGCAGGCGCGCAAACCGGCAGTGCGGCAGATG 2791
     |:::      :::::::::|||||||||          :::|||     |
1224 rLysAspThrGlnAsnThrAlaGlyAlaThrGlnProAlaAlaAsnThrA 1241

2792 CGCCGCGCCGCCGTTCGCGCCGTTCCCTATTATCCGTTACGCCGCCAACT 2841
     ||      :::   ::::::::::::|||:::::::::|||   |||:::
1241 la......GluValAlaLysGlnAspLeuValAspLeuThrLysProAla 1255

2842 TCG.......................GCAGAATCCCGTTTCAACACGCT 2867
     :::                      |||::::::      :::|||
1256 ThrGlyAlaAlaGlyAsnGlyAlaAspAlaLysAlaProAspThrThrAl 1272

2868 GACGGTAAACGGCAAATTGAACGGTCAGGGAACATTCCGCTTTATGTCGG 2917
      ::::::   |||   |||:::|||   |||
1272 aAlaThrValGlyAspLeuArgGlyLeuGly..................... 1282

2918 AACTCTTCGGCTACCGCAGCGGCAAATTGAAGCTGGCGGAAAGTTCCGAA 2967
                    |||:::|||            |||::::::::::
1283 ........TrpValLeuSerAlaLysLysThrAlaAspGluThrGlnAsp 1296

2968 GGCACTTACACCTTGGCTGTCAACAATACCGGCAAC...GAACCCGTAAG 3014
        :::     ||||||:::|||:::::::::   |||   |||::
1297 LysGluPheHisAlaAlaValLysAsnAlaAsnGluValGluPheValGl 1313

3015 TCTCGAGCAATTGACGGTAGTGGAAGGAAAAGACAACACACCGCTGTCCG 3064
     :  :::     ||||||            ||||||
1313 yLysAsnGlyAlaThrValSerAlaLysThrAspAsn.............. 1325

3065 AAAATCTTAATTTCACCCTGCAAAACGAACACGTCGATGCCGGCGCATGG 3114

1325 ..................................................... 1325

3115 CGTTATCAGCTTATCCGCAAAGACGGCGAGTTCCGCCTGCATAATCCGGT 3164
                        :::|||:::     :::   ||
1326 ....................AsnGlyLysHisThrValThrIleAspVa 1335

3165 CAAAGAACAAGAGCTTTCCGACAAACTCGGCAAGGCGGGAGAAACAGAGG 3214
     |   |||   ::::::::::|||   |||   |||      :::|||::::
1335 lAlaGluAlaLysValGlyAspGlyLeuGluLys......AspThrAspG 1350

3215 CCGCCTTGACGGCAAAA................................ 3231
     ::   :::      |||
1350 lyLysIleLysLeuLysValAspAsnThrAspGlyAsnAsnLeuLeuThr 1366

3232 ...CAGGCACAACTTGCCGCCAAACAACAGGCGGAAAAAGACAACGCGCA 3278
        :::|||      :::|||:::            :::   ||||||
1367 ValAspAlaThrLysGlyAlaSerValAlaLysGlyGluPheAsnAlaVa 1383

3279 AAGCCTTGACGCGCTGATTGCGGCCGGGCGCAATGCCACCGAAAAGGCAG 3328
      :::   ||||||      |||   |||   |||||||:::|||:::::::
1383 lThrThrAspAlaThrThrAlaGlnGlyThrAsnAlaAsnGluArgGlyL 1400

3329 AAAGTGTTGCCGAACCGGCCCGGCAGGCAGGCGGGGAAAATGCCGGCATT 3378
     ::   |||::::::                 |||:::|||   |||
1400 ysValValValLys.................GlySerAsn...GlyAla 1409

3379 ATGCAGGCGGAGGAAGAGAAAAAACGGGTGCAGGCGGATAAAGACACCGC 3428
        :::|||   :::|||||||:::|||   :::      |||:::||
1410 ThrAlaThrGluThrAspLysLysLysValAlaThrValGlyAspValAl 1426

3429 CTTGGCGAAACAGCGCGAAGCG..............GAAACCCGGCCGG 3463
     |   |||   ::       |||                |||:::      :
1426 aLysAlaIleAsnAspAlaAlaThrPheValLysValGluAsnAspAspS 1443

3464 CTACCACCGCCTTCCCCCGCGCCCGCCGCGCCCGCCGGGATTTGCCGCAA 3513
```

```
     :::::|||                                |||   |||
1443 erAlaThrIle.....................AspAspSerProThr 1451

3514 CCGCAGCCCCAACCGCAACCCCAACCGCAGCGCGACCTGATCAGCCGTTA 3563
       :::       :::                |||   ::::::
1452 AspAspGlyAlaAsnAspAlaLeuLysAlaGlyAspThrLeuThrLeuLy 1468

3564 TGCCAATAGCGGTTTGAGTGAATTTTCCGCCACGCTCAACAGCGTTTTCG 3613
      |||::::::::::|||:::                  |||   :::|||||
1468 sAlaGlyLysAsnLeuLysValLysArgAspGlyLysAsnIleThrPheA 1485

3614 CCGTACAGGACGAATTGGACCGCGTGTTTGCCGAAGACCGCCGCAACGCC 3663
     ||:::   ::::::|||                              :::
1485 laLeuAlaAsnAspLeu.............................Ser 1491

3664 GTTTGGACAAGCGGCATCCGGGACACCAAACACTACCGTTCGCAAGATTT 3713
     |||   ::::::   :::   |||
1492 ValLysSerAlaThrValSerAsp........................... 1499

3714 CCGCGCCTACCGCCAACAAACCGACCTGCGCCAAATCGGTATGCAGAAAA 3763
                        :::   :::|||          |||   :::   |
1500 ................LysLeuSerLeuGlyThrAsnGlyAsnLysValA 1511

3764 ACCTCGGCAGCGGGCGCGTCGGCATCCTGTTTTCG...CACAACCGGACC 3810
     ||:::   |||         |||:::   |||:::      ::::::|||
1511 snIleThrSerAspThrLysGlyLeuAsnPheAlaLysAspSerLysThr 1527

3811 GGA.........AACACCTTCGACGACGGCATCGGCAACTCGGCACGGCT 3851
     |||         |||         :::|||||||:::::::::
1528 GlyAspAspAlaAsnIleHisLeuAsnGlyIleAlaSerThrLeuThr.. 1543

3852 TGCCCACGGTGCCGTTTTCGGGCAATACGGCATCGGCAGGTTCGACATCG 3901
                                                   |||
1544 ..................................................AspThrL 1546

3902 GCATCAGCGCGGGCGCGGGTTTTAGTAGCGGCAGCCTTTCAGACGGCATC 3951
       ::::::::::|||||||      :::   |||      ::::::|||||||
1546 euLeuAsnSerGlyAlaThrThrAsnLeuGly......GlyAsnGlyIle 1560

3952 AGAGGCAAAATCCGCCGCCGCCGTGCTGCATTACGGCATTCAGGCAAGATA 4001
           :::   :::::::|||:::      .   :::::::::
1561 ThrAspAsnGluLysLysArgAlaAla.....SerValLysAspValLe 1575

4002 CCGCGCAGGTTTCGGCGGATTCGGCATCGAACCGCACATCGGCGCAACGC 4051
      :::||||||::::::      |||::::::|||
1575 uAsnAlaGlyTrpAsnValArgGlyValLysPro................ 1586

4052 GCTATTTCGTCCAAAAAGCGGATTACCGATACGAAAACGTCAATATCGCC 4101
                  :::|||:::   :::   ||||||:::::::::::::
1587 ..........AlaSerAlaAsnAsnGlnValGluAsnIleAspPheVal 1599

4102 ACCCCGGGCCTTGCATTCAACCGCTACCGCGCGGGCATTAAGGCAGATTA 4151
     :::
1600 Ala..................................................... 1600

4152 TTCATTCAAACCGGCGCAACACATTTCCATCACGCCTTATTTGAGCCTGT 4201
                                                     :::   :
1601 .....................................ThrTyrA 1603

4202 CCTATACCGATGCCGCTTCCGGCAAAGTCCGAACGCGCGTCAATACCGCC 4251
     ::   :::|||   :::|||||||          |||   ::::::::::::
1603 spThrValAspPheValSerGlyAspLysAspThr...ThrSerValThr 1618

4252 GTATTGGCGCAGGATTTCGGCAAAACCCGCAGTGCGGAATGGGGCGTAAA 4301
     |||   :::::::|||   ||||||      :::::::::   |||:::::
1619 ValGluSerLysAspAsnGlyLysArgThrGluValLysIleGlyAlaLy 1635

4302 CGCCGAAATC.................AAAGGTTTCACGCTGTCCCTCC 4333
     ::::::::::                |||   ||||||
1635 sThrSerValIleLysAspHisAsnGlyLysLeuPheThr.......... 1648

4334 ACGCTGCCGCCGCCAAGGGGCCGCAATTGGAAGCGCAGCACAGCGCGGGC 4383
                  |||   :::|||:::      ::::::   |||
```

```
    1649 .................GlyLysGluLeuLysAspAlaAsnAsnAsnGly 1659

    4384 ATC 4386
         :::
    1660 Val 1660

seq_name: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-09-268-347-47

seq_documentation_block:
; Sequence 47, Application US/09268347
; Patent No. 6335182
; GENERAL INFORMATION:
;  APPLICANT: Loosmore, Sheena M.
;  TITLE OF INVENTION: RECOMBINANT HAEMOPHILUS INFLUENZAE ADHESIN PROTEINS
;  FILE REFERENCE: 1038-860
;  CURRENT APPLICATION NUMBER: US/09/268,347
;  CURRENT FILING DATE: 1999-03-16
;  NUMBER OF SEQ ID NOS: 54
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 47
;   LENGTH: 2354
;   TYPE: PRT
;   ORGANISM: Haemophilus influenzae
US-09-268-347-47


alignment_scores:
            Quality:  254.50              Length:   1664
              Ratio:  0.346                 Gaps:     75
 Percent Similarity:  44.231   Percent Identity: 19.591

alignment_block:
 US-09-303-518D-653 x US-09-268-347-47   ..

 Align seg 1/1  to: US-09-268-347-47  from: 1  to: 2354

       7 ACAACCGACAAACGGACAACCGAA.........ACACACCGCAAAGCCCC 47
         |||   ||||||:::|||   :::         |||   ::::::
     246 ThrGlyAspLysAsnThrLeuAspValValLeuThrAlaLysGluAsnGl 262

      48 TAAAACCGGCCGCATCCGCTTCTCGCCCGCTTACTTAGCCATATGCCTGT 97
         ||||||   ::::::::::|||:::|||          :::|||      :
     262 yLysThrThrGluValLysPheThrProLysThrSerValIleLysGluL 279

      98 CGTTCGGCATTCTGCCCCAAGCCCGGGCGGGACACACTTATTTCGGCATC 147
         ::   |||   |||       ::::::   ::::::            :::
     279 ysAspGlyLysLeuPheThrGlyLysGluAsnAsnAspThrAsnLysVal 295

     148 AACTACCAATACTATCGCGACTTTGCCGAAAATAAAGGCAAGTTTGCAGT 197
         :::   :::          |||   :::::::::   :::   ::::::::
     296 ThrSerAsnThrAlaThrAspAsnThrAspGluGlyAsnGlyLeuValTh 312

     198 CGGGGCGAAAGATATTGAGGTTTACAACAAAAAAGGGGAGTTGGTCGGCA 247
         ::::          |||::::::   ||||||   |||       |||
     312 rAlaLysAlaValIleAspAlaValAsnLysAlaGlyTrpArgValLysT 329

     248 AATCGATGACGAAAGCCCCGATGATTGATTTT...TCTGTGGTATCGCGT 294
           :::   :::::::::          ||||||   :::|||:::|||
     329 hrThrThrAlaAsnGlyGlnAsnGlyAspPheAlaThrValAlaSerGly 345

     295 AACGGCGTGGCGGCATTGGCGGGCGATCAATATATTGTGAGCGTGGCACA 344
         :::::::|||:::      :::|||||||          :::|||||||:::
     346 ThrAsnValThrPheGluSerGlyAspGlyThrThrAlaSerValThrLy 362

     345 TAACGGCGGCTATAACAATGTTGATTTTGGTGCGGAGGGAAGCAATCCCG 394
         :::   :::   |||:::
     362 sAspThrAsnGlyAsnGly............................... 368

     395 ATCAGCACCGCTTTTCTTACCAAATTGTGAAAAGAAATAATTATAAAGCA 444
                               |||:::   :::   :::   |||:::
     369 .....................IleThrValLysTyrAspAlaLysVal 377

     445 GGGACTAACGGCCATCCTTATGGCGGCGATTATCATATGCCGCGTTTGCA 494

         |||   :::|||      :::   :::|||
     378 Gly...AspGlyLeuLysPheAspSerAsp.................Ly 387

     495 CAAATTTGTCACAGATGCAGAACCTGTTGAGATGACCAGTTATATGGATG 544
         |||:::|||:::|||:::      :::   :::|||           |
     387 sLysIleValAlaAspThrThrAlaLeuThrValThr............G 400

     545 GGTGGAAATACGCTGATTTAAATAAATACCCTGATCGTGTTCGAATC... 591
         ||   |||   |||::::::   |||      |||:::   ::::::
     400 lyGlyLysValAlaGluIleAlaLysGluAspAspLysLysLysLeuVal 416

     592 .....................GGAGCAGGCAGACAATATTGGCGG..... 615
                                  :::   |||:::      |||:::
     417 AsnAlaGlyAspLeuValThrAlaLeuGlyAsnLeuSerTrpLysAlaLy 433

     616 .......TCTGATGAAGAC.........GAACCCAATAACCGCGAAAGTT 649
                 :::|||   |||         |||      ::::::::::::
     433 sAlaGluAlaAspThrAspGlyAlaLeuGluGlyIleSerLysAspGlnG 450

     650 CATATCATATTGCAAGCGCATATTCTTGGCTCGTCGGTGGCAAT...ACC 696
         ::          :::::::::   ::::::   :::|||   |||
     450 luValLysAlaGlyGluThrValThrPheLysAlaGlyLysAsnLeuLys 466

     697 TTTGCACAAAATGGATCAGGT............................. 717
               |||:::|||::::::
     467 ValLysGlnAspGlyAlaAsnPheThrTyrSerLeuGlnAspAlaLeuTh 483

     718 .....................GGTGGCACAGTCAACTTAGGTAGCGAAA 745
                                 |||||||||:::|||   |||::::::
     483 rGlyLeuThrSerIleThrLeuGlyGlyThrThrAsnGlyGlyAsnAspA 500

     746 AA.........ATTAAACATAGCCCATATGGTTTTTTACCAACAGGAGGC 786
                 |||:::   :::         :::   |||:::|||:::
     500 laLysThrValIleAsnLysAspGlyLeuThrIleThrProAlaGlyAsn 516

     787 TCATTTGGCGACAGTGGCTCACCAATGTTTATCTATGATGCCCAAAAGCA 836
         :::   |||   :::|||:::      :::
     517 Gly...GlyThrThrGlyThrAsnThrIleSer.................. 526

     837 AAAGTGGTTAATTAATGGGGTATTGCAAACAGGCAACCCCTATATA...G 883
               :::   :::      :::::::::||||||      |||   :
     527 .......ValThrLysAspGlyIleLysAlaGlyAsnLysAlaIleThrA 541

     884 GAAAAAGCAATGGCTTCCAGCTAGTTCGTAAAGATTGGTTCTATGATGAA 933
         ::   ::::::|||::::::                        ||||||:::
     541 snValAlaSerGlyLeuArgAla..................TyrAspAsp 551

     934 ATCTTTGCTGGAGATACCCATTCAGTATTCTACGAACCACATCAAAATGG 983

     552 Ala.................................................. 552

     984 GAAATACTTTTTTAACGACAATAATAATGGCGCAGGAAAAATCGATGCCA 1033
               |||:::      ||||||:::|||      :::
     553 .......AsnPheAspValLeuAsnAsnSerAlaThrAspLeu....... 564

    1034 AACATAAACACTATTCTCTACCTTATAGATTAAAAACACGAACCGTTCAA 1083
           ::::::|||   :::      |||:::
     565 ..AsnArgHisValGluAspAlaTyrLys.................Gly 574

    1084 TTGTTTAATGTTTCTTTATCCGAGACAGCAAGAGAACCTGTTTAT..... 1128
         |||:::|||::::::   :::::::::   ::::::|||:::
     575 LeuLeuAsnLeuAsnGluLysAsnAlaAsnLysGlnProLeuValThrAs 591

    1129 .CATGCTGCAGGTGGGGTCAACAGTTATCGACCCAGACTGAATAATGGAG 1177
            :::|||:::   |||::::::   |||
     591 pSerThrAlaAlaThrValGlyAspLeuArgLys.............. 602

    1178 AAAATATTTCCTTTATTGACAAAGGAAAAGGTGAATTGATACTTACCAGC 1227
                                        |||   ::::::   |||:::
     603 .......................LeuGlyTrpValValSerThrLys 610

    1228 AACATCAACCAAGGCGCGGGCGGTTTGTATTTTGAGGGTAATTTTACGGT 1277
         |||                  |||         |||:::|||      ||
```

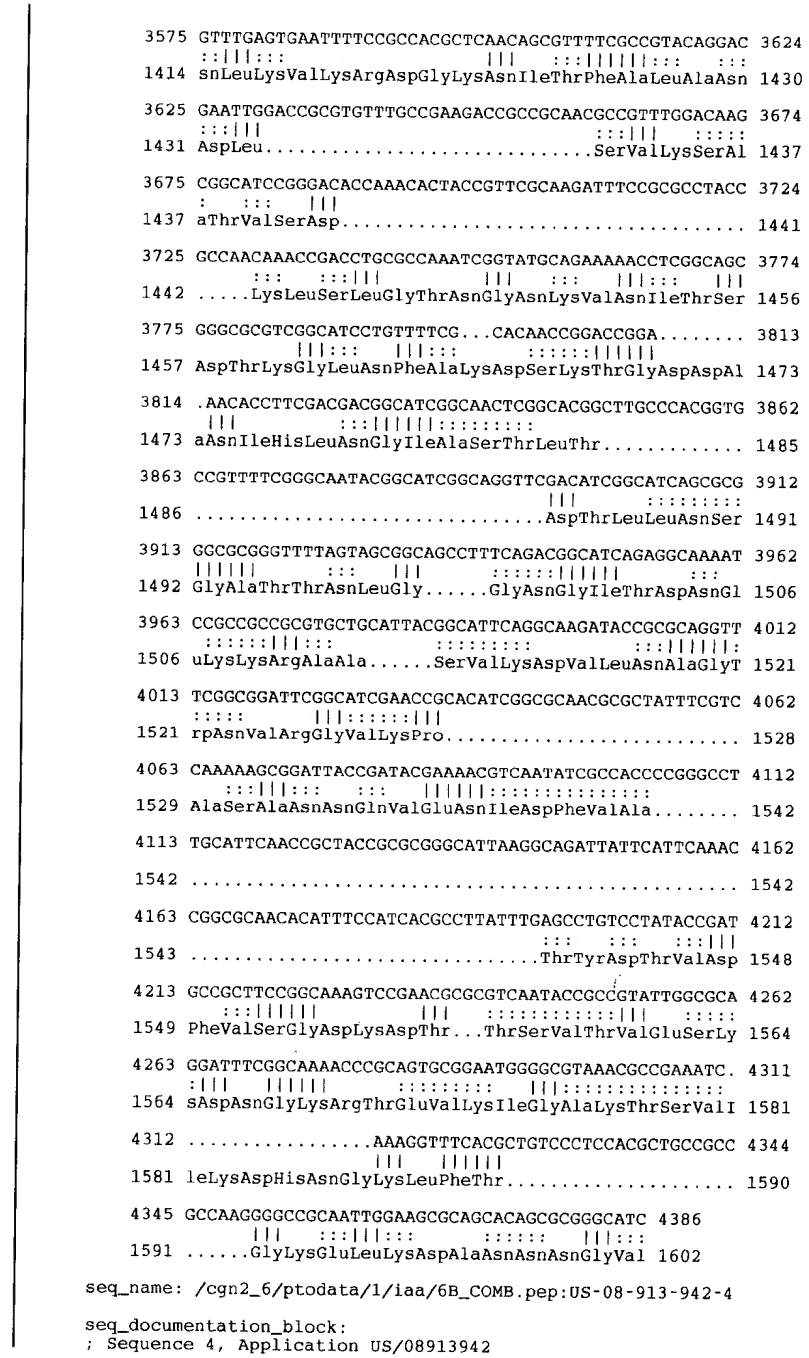
```
 611 Asn.................GlyThrLysGluGluSerAsn...GlnVa 620

1278 CTCGCCTAAAAACAACGAAACGTGGCAAGGCGCGGGCGTTCATATCAGTG 1327
     |:::      ::::::      :::   |||||||||
 620 lLysGlnAlaAspGluValLeuPheThrGlyAlaGly............. 632

1328 ATGGCAGTACCGTTACTTGGAAAGTAAACGGCGTGGCAAACGACCGCCTG 1377
       ::::::|||||||||   |||
 633 ..AlaAlaThrValThrSerLys........................... 639

1378 TCCAAAATCGGCAAAGGCACGCTGCTGGTTCAAGCCAAAGGGGAAAACCA 1427
     |||:::   ||||||   |||:::
 640 SerGluAsnGlyLysHisThrIle.......................... 647

1428 AGGCTCGGTCAGCGTGGGCGACGGTAAAGTCATCTTAGATCAGCAGGCGG 1477
       :::||||||||||:::::   |||
 648 ....ThrValSerValAlaGluThrLys..................... 655

1478 ACGATCAAGGCAAAAAACAAGCCTTTAGTGAAATCGGCTTGGTCAGCGGC 1527
                           ::::::   ||||||   :::
 656 .........................AlaAspCysGlyLeuGluLysAsp 663

1528 AGGGGGACGGTGCAACTGAATGCCGATAATCAGTTCAACCCCGACAAACT 1577
          |||::::::|||::::::|||||||||   :::   :::   ||
 664 GlyAspThrIleLysLeuLysValAspAsnGlnAsnThrAspAsnValLe 680

1578 CTATTTCGGCTTTCGCGGCGGACGTTTGGATTTGAACGGGCATTCGCTTT 1627
     |    |||   :::|||      :::      :::|||   :::
 680 uThrValGlyAsnAsnGlyThrAlaValThrLysGlyGlyPheGluThrV 697

1628 CGTTCCACCGCATTCAAAATACCGATGAAGGGGCGATGATTGTCAAC... 1674
                       ::::::|||:::|||   :::   |||:::
 697 alLysThrGlyAlaThrAspAlaAspArgGlyLysValThrValLysAsp 713

1675 .........CACAATCAAGACAAAGAA..................... 1692
             ::::::   ||||||:::
 714 AlaThrAlaAsnAspAlaAspLysLysValAlaThrValLysAspValAl 730

1693 ............TCCACCGTTACCATTACAGGCAATAAAGATATTACTA 1729
                |||:::::::|||::::::   :::::::::::::||||
 730 aThrAlaIleAsnSerAlaAlaThrPheValLysThrGluAsnLeuThrT 747

1730 CAACC...GGCAATAACAACAACTTGGATAGCAAAAAAGAAATTGCCTAC 1776
     ||:::      ::::::|||      |||:::   |||:::   |||
 747 hrSerIleAspGluAspAsnProThrAspAsnGlyLysAspAspAlaLeu 763

1777 AACGGTTGG.................TTTGGCGAGAAAGATGCAACCAA 1808
     ::::::                        |||::::::      :::||
 764 LysAlaGlyAspThrLeuThrPheLysAlaGlyLysAsnLeuLysValLy 780

1809 AACGAACGGGCGG.....................CTCAATCTGAATTACC 1837
     |   :::|||:::                      ||||||:::   :
 780 sArgAspGlyLysAsnIleThrPheAspLeuAlaLysAsnLeuGluValL 797

1838 AACCGGAAGAAGCGGATCGCACTTTACTGCTTTCCGGCGGAACAAATTTA 1887
     ::      :::::::::   ||||||   ::::::|||:::|||
 797 ysThrAlaLysValSerAspThrLeuThrIleGlyGlyAsnThrProThr 813

1888 AACGGC.....................AATATCACGCAAACAAACGGCAA 1916
     :::|||                     |||||||||:::|||
 814 GlyGlyThrThrAlaThrProLysValAsnIleThrSerThrAlaAspGl 830

1917 ACTGTTTTTCAGCGGCAGACCGACACCGCACGCCTACAATCATTTAGGAA 1966
       |||   |||:::   :::   |||       |||              ||||
 830 yLeuAsnPheAla...LysGluThrAlaAspAla.........SerGlyS 843

1967 GCGGGTGGTCAAAAATGGAAGGTATCCCACAAGGAGAAATCGTGTGGGAC 2016
     ||           ::::::||||||
 843 erLysAsnValTyrLeuLysGlyIleAla..................... 852

2017 AACGATTGGATCGACCGCACATTTAAAGCGGAAAACTTCCATATTCAG.. 2064
     :::         :::   :::      |||::::::   |||::::::
 853 ThrThrLeuThrGluProSerAlaGlyAlaLysSerSerHisValAspLe 869

2065 .GGCGGACAAGCGGTGGTTTCCCGCAATGTTGCCAAAGTGGAA....... 2106
       :::   :::|||:::   :::   |||:::|||:::::|||
 869 uAsnValAspAlaThrLysLysSerAsnAlaAlaSerIleGluAspValL 886

2107 .....GGCGATTGGCATTTAAGCAATCACGCCCAAGCAGTTTTCGGTGTC 2151
          :::   |||:::::::::::::::::::::::   |||      |||
 886 euArgAlaGlyTrpAsnIleGlnGlyAsnGlyAsnAsnValAspTyrVal 902

2152 GCACCGCATCAAAGCCACACAATCTGTACACGTTCGGACTGGACGGGTCT 2201
     |||   ::::::      |||:::      :::|||   ||||||
 903 AlaThrTyrAsp......ThrValAsnPheThrAspAspSerThrGlyTh 917

2202 GACAAGTTGTACCGAAAAAACCATTACCGACGATAAA......GTGATTG 2245
       |||:::   |||            :::|||   |||         :::
 917 rThrThrValThrValThrGlnLysAlaAspGlyLysGlyAlaAspValL 934

2246 CTTCATTGAGCAAGACCGACATC........................... 2268
            :::||||||:::::
 934 ysIleGlyAlaLysThrSerValIleLysAspHisAsnGlyLysLeuPhe 950

2269 ..........................AGAGGCAATGTCAGCCTTGCCGA 2291
                                  :::|||   :::         ||
 951 ThrGlyLysAspLeuLysAspAlaAsnAsnGlyAlaThrValSerGluAs 967

2292 TCACGCTCATTTAAATCTCACAGGACTTGCCACACTCAACGGCAATCTTA 2341
     |   :::      :::   |||||||||:::|||   :::      ::::
 967 pAspGlyLysAspThrGlyThrGlyLeuValThrAlaLysThrValIleA 984

2342 GTGCAGGCGGAGACACGCACTATACGGTTACG...CGCAACGCCACCCAA 2388
     :::||   :::   :::   :::   ||||||   ::::::||||||
 984 spAlaValAsnLysSerGlyTrpArgValThrGlyGluGlyAlaThrAla 1000

2389 AACGGCAACCTCAGCCTCGTG......GGCAATGCCCAAGCAACATTTAA 2432
     :::   :::   :::   |||      ||||||||:::::::::   ::
1001 GluThrGlyAlaThrAlaValAsnAlaGlyAsnAlaGluThrValThrSe 1017

2433 TCAAGCCACATTA.........AACGGCAAC............ACATCGG 2461
     :   :::::::::         |||||||||           |||   :
1017 rGlyThrSerValAsnPheLysAsnGlyAsnAlaThrThrAlaThrValS 1034

2462 CTTCGGACAATGCTTCATTTAATCTAAGCAACAACGCCGTACAAAACGGC 2511
     :::::||||||::::::::::|||:::::::   ::::::      :::
1034 erLysAspAsnGlyAsnIleAsnValLysTyrAspValAsnValGlyAsp 1050

2512 AGTCTGACGCTTTCCGACAACGCTAAGGCAAACGTAAGCCATTCCGCACT 2561
     :::|||   :::::::|||:::   |||      ::::::   ::::::||
1051 GlyLeuLysIleGlyAspAspLysLysIleValAlaAspThrThrThrLe 1067

2562 C.........AACGGCAATGTCTCCCTA...................... 2580
     .   |         :::|||:::||||||:::
1067 uThrValThrGlyGlyLysValSerValProAlaGlyAlaAsnSerValA 1084

2581 ......................GCCGATAAGGCAGTATTCCATTTTGAA 2607
                            |||:::      :::      ::::::
1084 snAsnAsnLysLysLeuValAsnAlaGluGlyLeuAlaThrAlaLeuAsn 1100

2608 AACAGCCGCTTTACCGGAAAA............................. 2628
     |||      :::|||:::|||
1101 AsnLeuSerTrpThrAlaLysAlaAspLysTyrAlaAspGlyGluSerGl 1117

2629 ..................ATCAGCGGCGGC.................... 2640
                         ::::::::::|||
1117 uGlyGluThrAspGlnGluValLysAlaGlyAspLysValThrPheLysA 1134

2641 ............................AAGGATACGGCATTACAC 2658
                                  ||||||   :::
1134 laGlyLysAsnLeuLysValLysGlnSerGluLysAspPheThrTyrSer 1150

2659 TTAAAAGACAGCGAATGGACGCTGCCGTCG............GGCACGGA 2696
     |||:::|||:::         |||   |||          ||||||
1151 LeuGlnAspThrLeuThrGlyLeuThrSerIleThrLeuGlyGlyThrAl 1167
```





seq_name: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-09-377-155-33

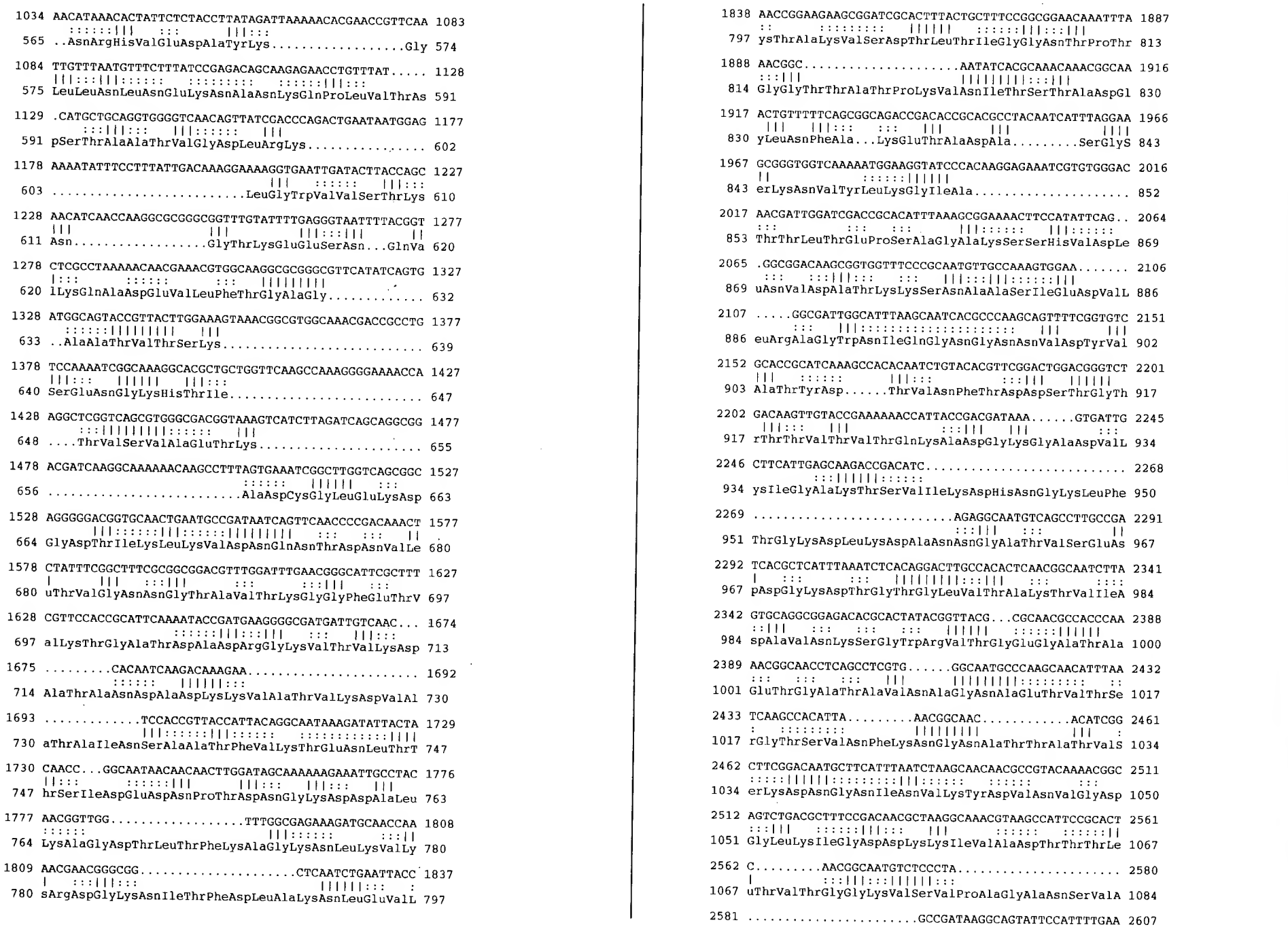
```
seq_documentation_block:
; Sequence 33, Application US/09377155
; Patent No. 6197312
; GENERAL INFORMATION:
;  APPLICANT:  PEAK, Ian Richard Anselm
;  APPLICANT:  JENNINGS, Michael Paul
;  APPLICANT:  MOXON, E. Richard
;  TITLE OF INVENTION:  NOVEL SURFACE ANTIGEN
;  FILE REFERENCE:  065064/0128
;  CURRENT APPLICATION NUMBER: US/09/377,155
;  CURRENT FILING DATE:  1999-08-19
;  PRIOR APPLICATION NUMBER:  PCT/AU98/01031
;  PRIOR FILING DATE:  1998-12-14
;  PRIOR APPLICATION NUMBER:  GB 9726398.2
;  PRIOR FILING DATE:  1997-12-12
;  NUMBER OF SEQ ID NOS:  33
;  SOFTWARE:  PatentIn Ver. 2.0
; SEQ ID NO 33
;   LENGTH: 2353
;   TYPE: PRT
;   ORGANISM: Haemophilus influenzae
US-09-377-155-33

alignment_scores:
            Quality:  253.50          Length:   1664
              Ratio:  0.344             Gaps:     75
 Percent Similarity:  44.231  Percent Identity: 19.591

alignment_block:
 US-09-303-518D-653 x US-09-377-155-33   ..

 Align seq 1/1  to: US-09-377-155-33  from: 1  to: 2353

       7 ACAACCGACAAACGGACAACCGAA.....ACACACCGCAAAGCCCC 47
         |||   ||||||:::|||   :::       |||  ::::::
     246 ThrGlyAspLysAsnThrLeuAspValValLeuThrAlaLysGluAsnGl 262

      48 TAAAACCGGCCGCATCCGCTTCTCGCCCGCTTACTTAGCCATATGCCTGT 97
          ||||||  ::::::::::|||:::|||       :::|||     :
     262 yLysThrThrGluValLysPheThrProLysThrSerValIleLysGluL 279

      98 CGTTCGGCATTCTGCCCCAAGCCCGGGCGGGACACACTTATTTCGGCATC 147
         ::  |||  |||     ::::::  ::::::           :::
     279 ysAspGlyLysLeuPheThrGlyLysGluAsnAsnAspThrAsnLysVal 295

     148 AACTACCAATACTATCGCGACTTTGCCGAAAATAAAGGCAAGTTTGCAGT 197
         :::  :::         |||  :::::::::  :::  ::::::::
     296 ThrSerAsnThrAlaThrAspAsnThrAspGluGlyAsnGlyLeuValTh 312

     198 CGGGGCGAAAGATATTGAGGTTTACAACAAAAAAGGGGAGTTGGTCGGCA 247
         ::::        |||::::::  ||||||  |||     |||
     312 rAlaLysAlaValIleAspAlaValAsnLysAlaGlyTrpArgValLysT 329

     248 AATCGATGACGAAAGCCCCGATGATTGATTTT...TCTGTGGTATCGCGT 294
         :::  :::::::::          ||||||  :::|||:::|||
     329 hrThrThrAlaAsnGlyGlnAsnGlyAspPheAlaThrValAlaSerGly 345

     295 AACGGCGTGGCGGCATTGGCGGGCGATCAATATATTGTGAGCGTGGCACA 344
         ::::::|||:::      :::||||||         :::|||||:::
     346 ThrAsnValThrPheGluSerGlyAspGlyThrThrAlaSerValThrLy 362

     345 TAACGGCGGCTATAACAATGTTGATTTTGGTGCGGAGGGAAGCAATCCCG 394
         :::  :::  |||:::
     362 sAspThrAsnGlyAsnGly.............................. 368

     395 ATCAGCACCGCTTTTCTTACCAAATTGTGAAAAGAAATAATTATAAAGCA 444
                        |||:::  :::  :::  |||:::
     369 ......................IleThrValLysTyrAspAlaLysVal 377

     445 GGGACTAACGGCCATCCTTATGGCGGCGATTATCATATGCCGCGTTTGCA 494
         |||  :::|||      :::  :::||||
     378 Gly...AspGlyLeuLysPheAspSerAsp.................Ly 387

     495 CAAATTTGTCACAGATGCAGAACCTGTTGAGATGACCAGTTATATGGATG 544
          |||:::|||:::|||:::      :::  :::|||          |
     387 sLysIleValAlaAspThrThrAlaLeuThrValThr...........G 400

     545 GGTGGAAATACGCTGATTTAAATAAATACCCTGATCGTGTTCGAATC... 591
         ||  |||  |||::::::  |||      |||:::  ::::::
     400 lyGlyLysValAlaGluIleAlaLysGluAspAspLysLysLysLeuVal 416

     592 .....GGAGCAGGCAGACAATATTGGCGG..... 615
                        :::  |||:::      |||:::
     417 AsnAlaGlyAspLeuValThrAlaLeuGlyAsnLeuSerTrpLysAlaLy 433

     616 .......TCTGATGAAGAC.........GAACCCAATAACCGCGAAAGTT 649
                ::|||  |||         |||      ::::::::::::
     433 sAlaGluAlaAspThrAspGlyAlaLeuGluGlyIleSerLysAspGlnG 450

     650 CATATCATATTGCAAGCGCATATTCTTGGCTCGTCGGTGGCAAT...ACC 696
         ::        :::::::::  ::::::  :::|||  |||
     450 luValLysAlaGlyGluThrValThrPheLysAlaGlyLysAsnLeuLys 466

     697 TTTGCACAAAATGGATCAGGT........................... 717
               |||:::|||::::::
     467 ValLysGlnAspGlyAlaAsnPheThrTyrSerLeuGlnAspAlaLeuTh 483

     718 .....GGTGGCACAGTCAACTTAGGTAGCGAAA 745
                |||||||||:::|||  |||::::::
     483 rGlyLeuThrSerIleThrLeuGlyGlyThrThrAsnGlyGlyAsnAspA 500

     746 AA.........ATTAAACATAGCCCATATGGTTTTTTACCAACAGGAGGC 786
                 |||:::  :::        :::  |||:::|||:::
     500 laLysThrValIleAsnLysAspGlyLeuThrIleThrProAlaGlyAsn 516

     787 TCATTTGGCGACAGTGGCTCACCAATGTTTATCTATGATGCCCAAAAGCA 836
         :::  |||  :::|||:::      :::
     517 Gly...GlyThrThrGlyThrAsnThrIleSer.................. 526

     837 AAAGTGGTTAATTAATGGGGTATTGCAAACAGGCAACCCCTATATA...G 883
             :::  :::      :::::::::||||||      |||  :
     527 .......ValThrLysAspGlyIleLysAlaGlyAsnLysAlaIleThrA 541

     884 GAAAAAGCAATGGCTTCCAGCTAGTTCGTAAAGATTGGTTCTATGATGAA 933
         ::  ::::::|||::::::                    ||||||:::
     541 snValAlaSerGlyLeuArgAla.................TyrAspAsp 551

     934 ATCTTTGCTGGAGATACCCATTCAGTATTCTACGAACCACATCAAAATGG 983

     552 Ala.......................................... 552

     984 GAAATACTTTTTTAACGACAATAATAATGGCGCAGGAAAAATCGATGCCA 1033
                |||:::      ||||||:::|||      :::
     553 .......AsnPheAspValLeuAsnAsnSerAlaThrAspLeu....... 564

    1034 AACATAAACACTATTCTCTACCTTATAGATTAAAAACACGAACCGTTCAA 1083
         :::::::|||  :::      |||:::
     565 ..AsnArgHisValGluAspAlaTyrLys.................Gly 574

    1084 TTGTTTAATGTTTCTTTATCCGAGACAGCAAGAGAACCTGTTTAT..... 1128
         |||:::|||::::::  :::::::::  ::::::|||:::
     575 LeuLeuAsnLeuAsnGluLysAsnAlaAsnLysGlnProLeuValThrAs 591

    1129 .CATGCTGCAGGTGGGGTCAACAGTTATCGACCCAGACTGAATAATGGAG 1177
            :::|||:::  |||::::::  |||
     591 pSerThrAlaAlaThrValGlyAspLeuArgLys.............. 602

    1178 AAAATATTTCCTTTATTGACAAAGGAAAAGGTGAATTGATACTTACCAGC 1227
                                  |||  ::::::  |||:::
     603 ......................LeuGlyTrpValValSerThrLys 610

    1228 AACATCAACCAAGGCGCGGGCGGTTTGTATTTTGAGGGTAATTTTACGGT 1277
         |||                |||          |||:::|||      ||
     611 Asn.................GlyThrLysGluGluSerAsn...GlnVa 620
```

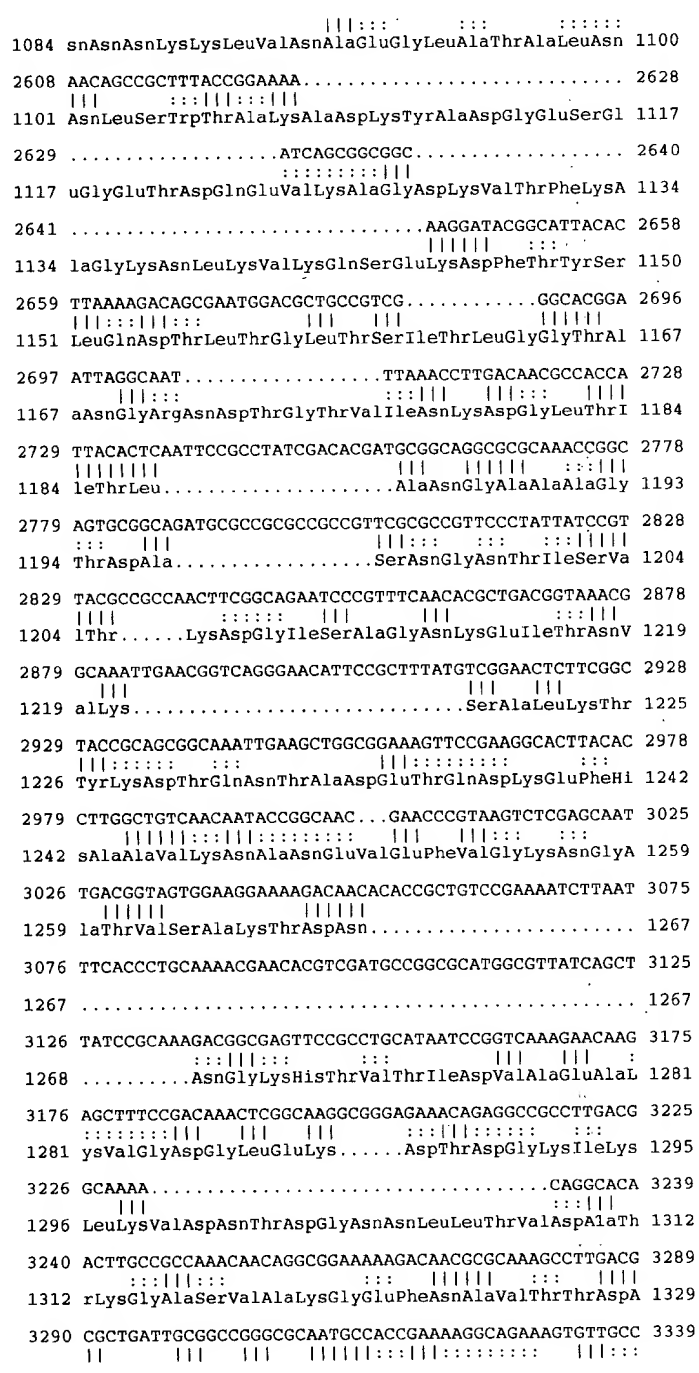


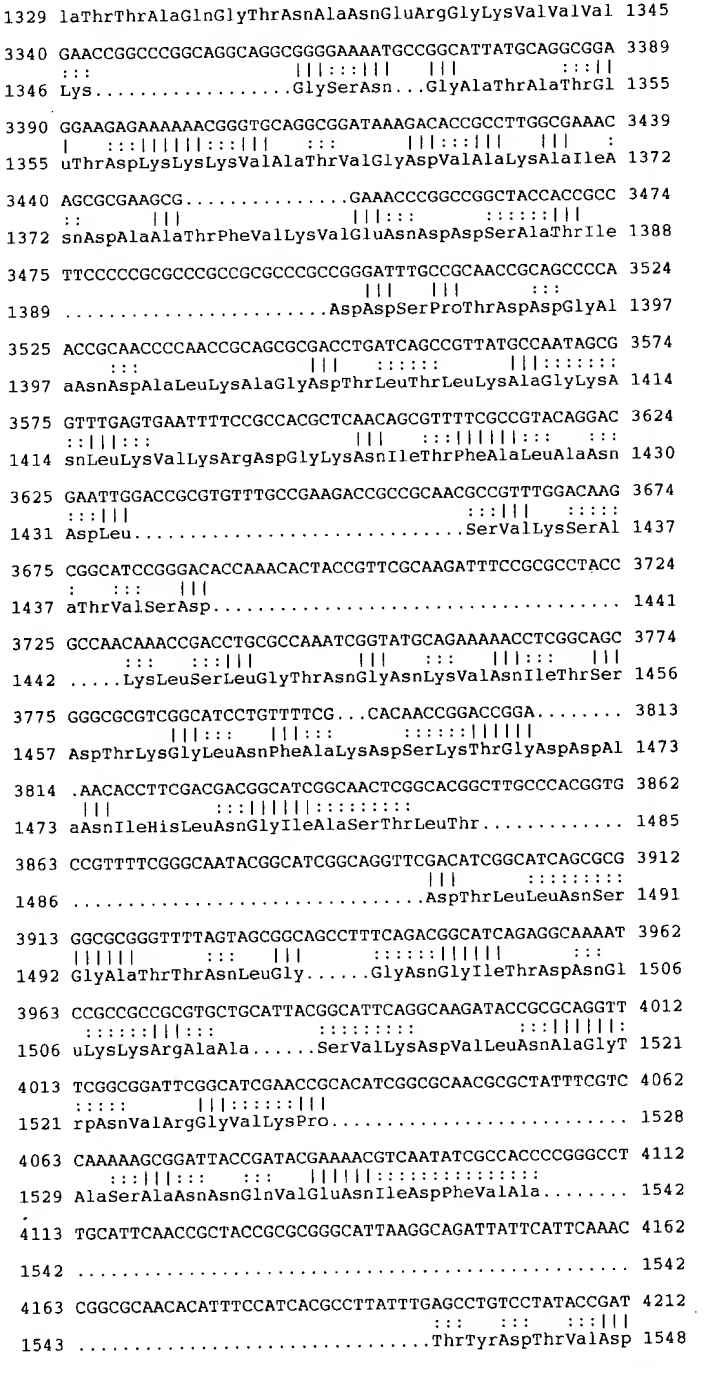
seq_name: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-08-913-942-4

seq_documentation_block:
; Sequence 4, Application US/08913942

```
; Patent No. 6200578
;  GENERAL INFORMATION:
;    APPLICANT:  St. Geme, Joseph
;    APPLICANT:  Barenkamp, Stephen J.
;    TITLE OF INVENTION:  HAEMOPHILUS ADHESION PROTEINS
;    NUMBER OF SEQUENCES:  19
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Flehr Hohbach Test Albritton & Herbert LLP
;      STREET:  Four Embarcadero Center, Suite 3400
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  United States
;      ZIP:  94111-4187
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/913,942
;      FILING DATE:  29-DEC-1997
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/409,995
;      FILING DATE:  24-MAR-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US96/4031
;      FILING DATE:  22-MAR-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Vance, Dolly A.
;      REGISTRATION NUMBER:  39,054
;      REFERENCE/DOCKET NUMBER:  A-61053-1/RFT/RMS/DAV
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 781-1989
;      TELEFAX:  (415) 398-3249
;      TELEX:  910 277299
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2353 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-913-942-4


alignment_scores:
            Quality:  253.50            Length:   1664
              Ratio:  0.344               Gaps:     75
 Percent Similarity:  44.231   Percent Identity: 19.591

alignment_block:
 US-09-303-518D-653 x US-08-913-942-4   ..

 Align seg 1/1  to: US-08-913-942-4  from: 1  to: 2353

       7 ACAACCGACAAACGGACAACCGAA.....ACACACCGCAAAGCCCC 47
         |||   ||||||:::|||   :::         |||   ::::::
     246 ThrGlyAspLysAsnThrLeuAspValValLeuThrAlaLysGluAsnGl 262

      48 TAAAACCGGCCGCATCCGCTTCTCGCCCGCTTACTTAGCCATATGCCTGT 97
          ||||||   ::::::::::|||:::|||          :::|||      :
     262 yLysThrThrGluValLysPheThrProLysThrSerValIleLysGluL 279

      98 CGTTCGGCATTCTGCCCCAAGCCCGGGCGGGACACACTTATTTCGGCATC 147
         ::   |||   |||      ::::::   ::::::              :::
     279 ysAspGlyLysLeuPheThrGlyLysGluAsnAsnAspThrAsnLysVal 295

     148 AACTACCAATACTATCGCGACTTTGCCGAAAATAAAGGCAAGTTTGCAGT 197
         :::   :::         |||   :::::::::   :::   ::::::::
     296 ThrSerAsnThrAlaThrAspAsnThrAspGluGlyAsnGlyLeuValTh 312

     198 CGGGGCGAAAGATATTGAGGTTTACAACAAAAAAGGGGAGTTGGTCGGCA 247
         ::::         |||:::::::   ||||||   |||      |||

     312 rAlaLysAlaValIleAspAlaValAsnLysAlaGlyTrpArgValLysT 329

     248 AATCGATGACGAAAGCCCCGATGATTGATTTT...TCTGTGGTATCGCGT 294
           :::   :::::::::         ||||||   :::|||:::|||
     329 hrThrThrAlaAsnGlyGlnAsnGlyAspPheAlaThrValAlaSerGly 345

     295 AACGGCGTGGCGGCATTGGCGGGCGATCAATATATTGTGAGCGTGGCACA 344
         ::::::|||:::      :::|||||||         :::|||||||:::
     346 ThrAsnValThrPheGluSerGlyAspGlyThrThrAlaSerValThrLy 362

     345 TAACGGCGGCTATAACAATGTTGATTTTGGTGCGGAGGGAAGCAATCCCG 394
          :::   :::   |||:::
     362 sAspThrAsnGlyAsnGly.............................. 368

     395 ATCAGCACCGCTTTTCTTACCAAATTGTGAAAAGAAATAATTATAAAGCA 444
                                 |||:::   :::   :::   |||:::
     369 .....IleThrValLysTyrAspAlaLysVal 377

     445 GGGACTAACGGCCATCCTTATGGCGGCGATTATCATATGCCGCGTTTGCA 494
         |||   :::|||      :::   :::|||
     378 Gly...AspGlyLeuLysPheAspSerAsp.....Ly 387

     495 CAAATTTGTCACAGATGCAGAACCTGTTGAGATGACCAGTTATATGGATG 544
          |||:::|||:::|||:::      :::   :::|||           |
     387 sLysIleValAlaAspThrThrAlaLeuThrValThr.....G 400

     545 GGTGGAAATACGCTGATTTAAATAAATACCCTGATCGTGTTCGAATC... 591
         ||   |||   |||::::::   |||      |||:::   ::::::
     400 lyGlyLysValAlaGluIleAlaLysGluAspAspLysLysLysLeuVal 416

     592 .....GGAGCAGGCAGACAATATTGGCGG..... 615
                                :::   |||:::      |||:::
     417 AsnAlaGlyAspLeuValThrAlaLeuGlyAsnLeuSerTrpLysAlaLy 433

     616 .......TCTGATGAAGAC.........GAACCCAATAACCGCGAAAGTT 649
                ::::|||   |||         |||      ::::::::::::
     433 sAlaGluAlaAspThrAspGlyAlaLeuGluGlyIleSerLysAspGlnG 450

     650 CATATCATATTGCAAGCGCATATTCTTGGCTCGTCGGTGGCAAT...ACC 696
         ::         :::::::::   ::::::   :::|||   |||
     450 luValLysAlaGlyGluThrValThrPheLysAlaGlyLysAsnLeuLys 466

     697 TTTGCACAAAATGGATCAGGT............................ 717
               |||:::|||::::::
     467 ValLysGlnAspGlyAlaAsnPheThrTyrSerLeuGlnAspAlaLeuTh 483

     718 .....GGTGGCACAGTCAACTTAGGTAGCGAAA 745
                                |||||||||:::|||   |||::::::
     483 rGlyLeuThrSerIleThrLeuGlyGlyThrThrAsnGlyGlyAsnAspA 500

     746 AA.........ATTAAACATAGCCCATATGGTTTTTTACCAACAGGAGGC 786
                    |||:::   :::         :::   |||:::|||:::
     500 laLysThrValIleAsnLysAspGlyLeuThrIleThrProAlaGlyAsn 516

     787 TCATTTGGCGACAGTGGCTCACCAATGTTTATCTATGATGCCCAAAAGCA 836
         :::   |||   :::|||:::      :::
     517 Gly...GlyThrThrGlyThrAsnThrIleSer.............. 526

     837 AAAGTGGTTAATTAATGGGGTATTGCAAACAGGCAACCCCTATATA...G 883
               :::   :::      :::::::::||||||      |||   :
     527 .......ValThrLysAspGlyIleLysAlaGlyAsnLysAlaIleThrA 541

     884 GAAAAAGCAATGGCTTCCAGCTAGTTCGTAAAGATTGGTTCTATGATGAA 933
         ::   :::::::|||::::::                    ||||||:::
     541 snValAlaSerGlyLeuArgAla.................TyrAspAsp 551

     934 ATCTTTGCTGGAGATACCCATTCAGTATTCTACGAACCACATCAAAATGG 983

     552 Ala........................................... 552

     984 GAAATACTTTTTTAACGACAATAATAATGGCGCAGGAAAAATCGATGCCA 1033
                    |||:::      ||||||:::|||      :::
     553 .......AsnPheAspValLeuAsnAsnSerAlaThrAspLeu..... 564
```



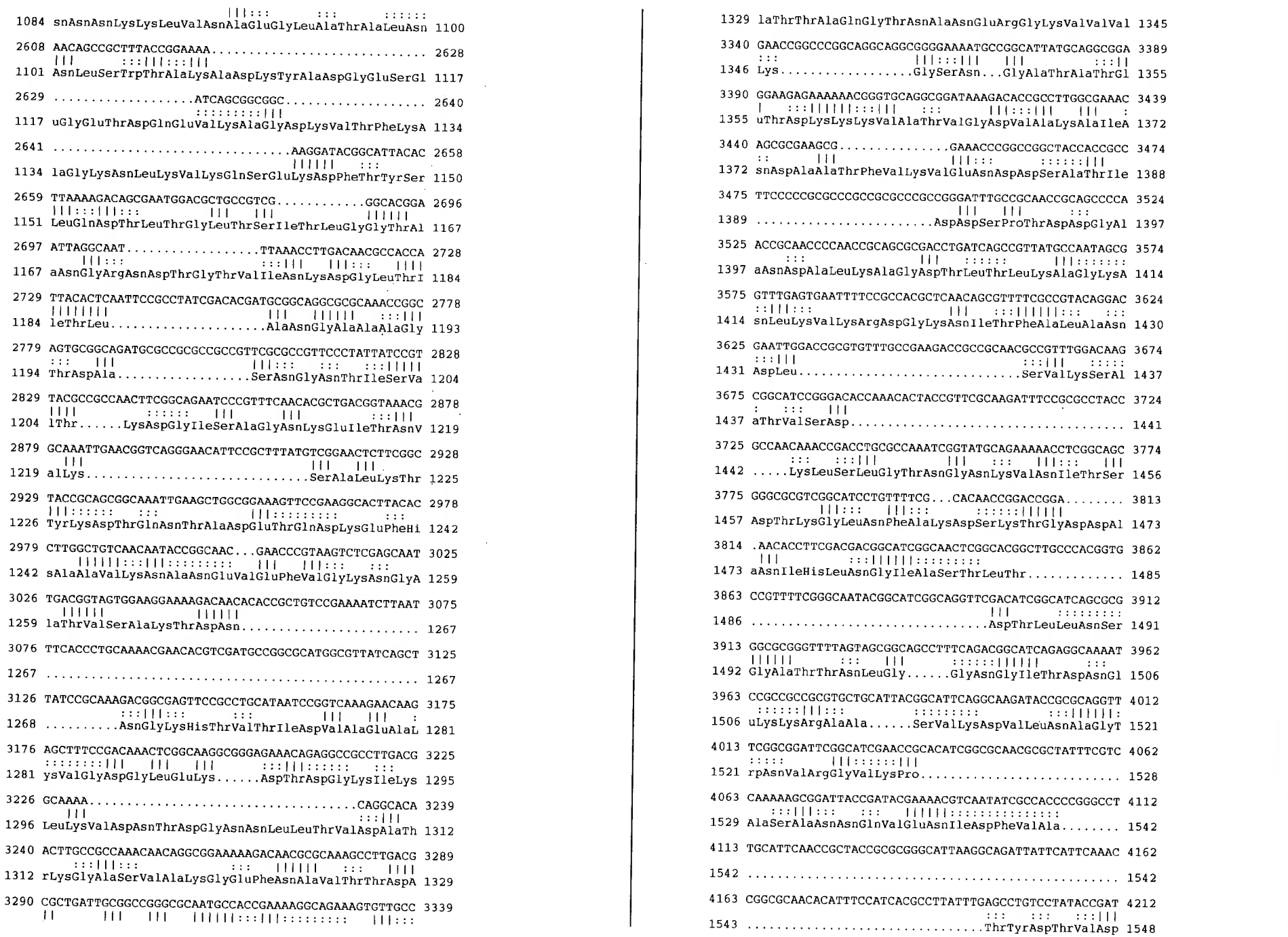

1084 snAsnAsnLysLysLeuValAsnAlaGluGlyLeuAlaThrAlaLeuAsn 1100
2608 AACAGCCGCTTTACCGGAAAA.... 2628
1101 AsnLeuSerTrpThrAlaLysAlaAspLysTyrAlaAspGlyGluSerGl 1117
2629 .....ATCAGCGGCGGC..... 2640
1117 uGlyGluThrAspGlnGluValLysAlaGlyAspLysValThrPheLysA 1134
2641 .....AAGGATACGGCATTACAC 2658
1134 laGlyLysAsnLeuLysValLysGlnSerGluLysAspPheThrTyrSer 1150
2659 TTAAAAGACAGCGAATGGACGCTGCCGTCG.....GGCACGGA 2696
1151 LeuGlnAspThrLeuThrGlyLeuThrSerIleThrLeuGlyGlyThrAl 1167
2697 ATTAGGCAAT.....TTAAACCTTGACAACGCCACCA 2728
1167 aAsnGlyArgAsnAspThrGlyThrValIleAsnLysAspGlyLeuThrI 1184
2729 TTACACTCAATTCCGCCTATCGACACGATGCGGCAGGCGCGCAAACCGGC 2778
1184 leThrLeu.....AlaAsnGlyAlaAlaAlaGly 1193
2779 AGTGCGGCAGATGCGCCGCGCCGCCGTTCGCGCCGTTCCCTATTATCCGT 2828
1194 ThrAspAla.....SerAsnGlyAsnThrIleSerVa 1204
2829 TACGCCGCCAACTTCGGCAGAATCCCGTTTCAACACGCTGACGGTAAACG 2878
1204 lThr.....LysAspGlyIleSerAlaGlyAsnLysGluIleThrAsnV 1219
2879 GCAAATTGAACGGTCAGGGAACATTCCGCTTTATGTCGGAACTCTTCGGC 2928
1219 alLys.....SerAlaLeuLysThr 1225
2929 TACCGCAGCGGCAAATTGAAGCTGGCGGAAAGTTCCGAAGGCACTTACAC 2978
1226 TyrLysAspThrGlnAsnThrAlaAspGluThrGlnAspLysGluPheHi 1242
2979 CTTGGCTGTCAACAATACCGGCAAC...GAACCCGTAAGTCTCGAGCAAT 3025
1242 sAlaAlaValLysAsnAlaAsnGluValGluPheValGlyLysAsnGlyA 1259
3026 TGACGGTAGTGGAAGGAAAAGACAACACACCGCTGTCCGAAAATCTTAAT 3075
1259 laThrValSerAlaLysThrAspAsn..... 1267
3076 TTCACCCTGCAAAACGAACACGTCGATGCCGGCGCATGGCGTTATCAGCT 3125
1267 1267
3126 TATCCGCAAAGACGGCGAGTTCCGCCTGCATAATCCGGTCAAAGAACAAG 3175
1268 .....AsnGlyLysHisThrValThrIleAspValAlaGluAlaL 1281
3176 AGCTTTCCGACAAACTCGGCAAGGCGGGAGAAACAGAGGCCGCCTTGACG 3225
1281 ysValGlyAspGlyLeuGluLys.....AspThrAspGlyLysIleLys 1295
3226 GCAAAA.....CAGGCACA 3239
1296 LeuLysValAspAsnThrAspGlyAsnAsnLeuLeuThrValAspAlaTh 1312
3240 ACTTGCCGCCAAACAACAGGCGGAAAAAGACAACGCGCAAAGCCTTGACG 3289
1312 rLysGlyAlaSerValAlaLysGlyGluPheAsnAlaValThrThrAspA 1329
3290 CGCTGATTGCGGCCGGGCGCAATGCCACCGAAAAGGCAGAAAGTGTTGCC 3339


1329 laThrThrAlaGlnGlyThrAsnAlaAsnGluArgGlyLysValValVal 1345
3340 GAACCGGCCCGGCAGGCAGGCGGGGAAAATGCCGGCATTATGCAGGCGGA 3389
1346 Lys.....GlySerAsn...GlyAlaThrAlaThrGl 1355
3390 GGAAGAGAAAAAACGGGTGCAGGCGGATAAAGACACCGCCTTGGCGAAAC 3439
1355 uThrAspLysLysLysValAlaThrValGlyAspValAlaLysAlaIleA 1372
3440 AGCGCGAAGCG.....GAAACCCGGCCGGCTACCACCGCC 3474
1372 snAspAlaAlaThrPheValLysValGluAsnAspAspSerAlaThrIle 1388
3475 TTCCCCCGCGCCCGCCGCGCCCGCCGGGATTTGCCGCAACCGCAGCCCCA 3524
1389 .....AspAspSerProThrAspAspGlyAl 1397
3525 ACCGCAACCCCAACCGCAGCGCGACCTGATCAGCCGTTATGCCAATAGCG 3574
1397 aAsnAspAlaLeuLysAlaGlyAspThrLeuThrLeuLysAlaGlyLysA 1414
3575 GTTTGAGTGAATTTTCCGCCACGCTCAACAGCGTTTTCGCCGTACAGGAC 3624
1414 snLeuLysValLysArgAspGlyLysAsnIleThrPheAlaLeuAlaAsn 1430
3625 GAATTGGACCGCGTGTTTGCCGAAGACCGCCGCAACGCCGTTTGGACAAG 3674
1431 AspLeu.....SerValLysSerAl 1437
3675 CGGCATCCGGGACACCAAACACTACCGTTCGCAAGATTTCCGCGCCTACC 3724
1437 aThrValSerAsp..... 1441
3725 GCCAACAAACCGACCTGCGCCAAATCGGTATGCAGAAAAACCTCGGCAGC 3774
1442 .....LysLeuSerLeuGlyThrAsnGlyAsnLysValAsnIleThrSer 1456
3775 GGGCGCGTCGGCATCCTGTTTTCG...CACAACCGGACCGGA..... 3813
1457 AspThrLysGlyLeuAsnPheAlaLysAspSerLysThrGlyAspAspAl 1473
3814 .AACACCTTCGACGACGGCATCGGCAACTCGGCACGGCTTGCCCACGGTG 3862
1473 aAsnIleHisLeuAsnGlyIleAlaSerThrLeuThr..... 1485
3863 CCGTTTTCGGGCAATACGGCATCGGCAGGTTCGACATCGGCATCAGCGCG 3912
1486 .....AspThrLeuLeuAsnSer 1491
3913 GGCGCGGGTTTTAGTAGCGGCAGCCTTTCAGACGGCATCAGAGGCAAAAT 3962
1492 GlyAlaThrThrAsnLeuGly.....GlyAsnGlyIleThrAspAsnGl 1506
3963 CCGCCGCCGCGTGCTGCATTACGGCATTCAGGCAAGATACCGCGCAGGTT 4012
1506 uLysLysArgAlaAla.....SerValLysAspValLeuAsnAlaGlyT 1521
4013 TCGGCGGATTCGGCATCGAACCGCACATCGGCGCAACGCGCTATTTCGTC 4062
1521 rpAsnValArgGlyValLysPro..... 1528
4063 CAAAAAGCGGATTACCGATACGAAAACGTCAATATCGCCACCCCGGGCCT 4112
1529 AlaSerAlaAsnAsnGlnValGluAsnIleAspPheValAla..... 1542
4113 TGCATTCAACCGCTACCGCGCGGGCATTAAGGCAGATTATTCATTCAAAC 4162
1542 1542
4163 CGGCGCAACACATTTCCATCACGCCTTATTTGAGCCTGTCCTATACCGAT 4212
1543 .....ThrTyrAspThrValAsp 1548

```
     4213 GCCGCTTCCGGCAAAGTCCGAACGCGCGTCAATACCGCCGTATTGGCGCA 4262
          :::||||||         |||   :::::::::::|||   :::::
     1549 PheValSerGlyAspLysAspThr...ThrSerValThrValGluSerLy 1564

     4263 GGATTTCGGCAAAACCCGCAGTGCGGAATGGGGCGTAAACGCCGAAATC. 4311
          :|||   ||||||      :::::::::   |||:::::::::::::::
     1564 sAspAsnGlyLysArgThrGluValLysIleGlyAlaLysThrSerValI 1581

     4312 .................AAAGGTTTCACGCTGTCCCTCCACGCTGCCGCC 4344
                          |||   ||||||
     1581 leLysAspHisAsnGlyLysLeuPheThr..................... 1590

     4345 GCCAAGGGGCCGCAATTGGAAGCGCAGCACAGCGCGGGCATC 4386
               |||   :::|||:::      ::::::   |||:::
     1591 ......GlyLysGluLeuLysAspAlaAsnAsnAsnGlyVal 1602

seq_name: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-09-669-974-33

seq_documentation_block:
; Sequence 33, Application US/09669974
; Patent No. 6333173
; GENERAL INFORMATION:
;  APPLICANT: PEAK, Ian Richard Anselm
;  APPLICANT:  JENNINGS, Michael Paul
;  APPLICANT:  MOXON, E. Richard
;  TITLE OF INVENTION: NOVEL SURFACE ANTIGEN
;  FILE REFERENCE: 065064/0128
;  CURRENT APPLICATION NUMBER: US/09/669,974
;  CURRENT FILING DATE:  2000-09-26
;  PRIOR APPLICATION NUMBER: US 09/377,155
;  PRIOR FILING DATE: 1999-08-19
;  PRIOR APPLICATION NUMBER: PCT/AU98/01031
;  PRIOR FILING DATE: 1998-12-14
;  PRIOR APPLICATION NUMBER: GB 9726398.2
;  PRIOR FILING DATE: 1997-12-12
;  NUMBER OF SEQ ID NOS: 33
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 33
;   LENGTH: 2353
;   TYPE: PRT
;   ORGANISM: Haemophilus influenzae
US-09-669-974-33


alignment_scores:
            Quality:  253.50              Length:   1664
              Ratio:  0.344                 Gaps:     75
 Percent Similarity:  44.231   Percent Identity: 19.591

alignment_block:
 US-09-303-518D-653 x US-09-669-974-33   ..

 Align seg 1/1  to: US-09-669-974-33  from: 1  to: 2353

       7 ACAACCGACAAACGGACAACCGAA.........ACACACCGCAAAGCCCC 47
         |||   ||||||:::|||   :::          |||   ::::::
     246 ThrGlyAspLysAsnThrLeuAspValValLeuThrAlaLysGluAsnGl 262

      48 TAAAACCGGCCGCATCCGCTTCTCGCCCGCTTACTTAGCCATATGCCTGT 97
         ||||||   ::::::::::|||:::|||          :::|||      :
     262 yLysThrThrGluValLysPheThrProLysThrSerValIleLysGluL 279

      98 CGTTCGGCATTCTGCCCCAAGCCCGGGCGGGACACACTTATTTCGGCATC 147
         ::   |||   |||      ::::::   ::::::                :::
     279 ysAspGlyLysLeuPheThrGlyLysGluAsnAsnAspThrAsnLysVal 295

     148 AACTACCAATACTATCGCGACTTTGCCGAAAATAAAGGCAAGTTTGCAGT 197
         :::   :::          |||   :::::::::   :::   ::::::::
     296 ThrSerAsnThrAlaThrAspAsnThrAspGluGlyAsnGlyLeuValTh 312

     198 CGGGGCGAAAGATATTGAGGTTTACAACAAAAAAGGGGAGTTGGTCGGCA 247
         ::::         |||::::::   ||||||   |||      |||
```

```
     312 rAlaLysAlaValIleAspAlaValAsnLysAlaGlyTrpArgValLysT 329

     248 AATCGATGACGAAAGCCCCGATGATTGATTTT...TCTGTGGTATCGCGT 294
           :::   :::::::::          ||||||   :::|||:::|||
     329 hrThrThrAlaAsnGlyGlnAsnGlyAspPheAlaThrValAlaSerGly 345

     295 AACGGCGTGGCGGCATTGGCGGGCGATCAATATATTGTGAGCGTGGCACA 344
         ::::::|||:::      :::||||||          :::||||||:::
     346 ThrAsnValThrPheGluSerGlyAspGlyThrThrAlaSerValThrLy 362

     345 TAACGGCGGCTATAACAATGTTGATTTTGGTGCGGAGGGAAGCAATCCCG 394
          :::   :::   |||:::
     362 sAspThrAsnGlyAsnGly.............................. 368

     395 ATCAGCACCGCTTTTCTTACCAAATTGTGAAAAGAAATAATTATAAAGCA 444
                               |||:::   :::   :::   |||:::
     369 ......................IleThrValLysTyrAspAlaLysVal 377

     445 GGGACTAACGGCCATCCTTATGGCGGCGATTATCATATGCCGCGTTTGCA 494
         |||   :::|||      :::   :::|||
     378 Gly...AspGlyLeuLysPheAspSerAsp.................Ly 387

     495 CAAATTTGTCACAGATGCAGAACCTGTTGAGATGACCAGTTATATGGATG 544
          |||:::|||:::|||:::      :::   :::|||            |
     387 sLysIleValAlaAspThrThrAlaLeuThrValThr............G 400

     545 GGTGGAAATACGCTGATTTAAATAAATACCCTGATCGTGTTCGAATC... 591
         ||   |||   |||::::::   |||      |||:::   ::::::
     400 lyGlyLysValAlaGluIleAlaLysGluAspAspLysLysLysLeuVal 416

     592 .....................GGAGCAGGCAGACAATATTGGCGG..... 615
                              :::   |||:::      |||:::
     417 AsnAlaGlyAspLeuValThrAlaLeuGlyAsnLeuSerTrpLysAlaLy 433

     616 .......TCTGATGAAGAC.........GAACCCAATAACCGCGAAAGTT 649
               :::|||   |||          |||      ::::::::::::
     433 sAlaGluAlaAspThrAspGlyAlaLeuGluGlyIleSerLysAspGlnG 450

     650 CATATCATATTGCAAGCGCATATTCTTGGCTCGTCGGTGGCAAT...ACC 696
         ::         :::::::::   ::::::   :::|||   |||
     450 luValLysAlaGlyGluThrValThrPheLysAlaGlyLysAsnLeuLys 466

     697 TTTGCACAAAATGGATCAGGT............................. 717
               |||:::|||::::::
     467 ValLysGlnAspGlyAlaAsnPheThrTyrSerLeuGlnAspAlaLeuTh 483

     718 .....................GGTGGCACAGTCAACTTAGGTAGCGAAA 745
                              |||||||||:::|||   |||::::::
     483 rGlyLeuThrSerIleThrLeuGlyGlyThrThrAsnGlyGlyAsnAspA 500

     746 AA.........ATTAAACATAGCCCATATGGTTTTTTACCAACAGGAGGC 786
                   |||:::   :::          :::   |||:::|||:::
     500 laLysThrValIleAsnLysAspGlyLeuThrIleThrProAlaGlyAsn 516

     787 TCATTTGGCGACAGTGGCTCACCAATGTTTATCTATGATGCCCAAAAGCA 836
         :::   |||   :::|||:::      :::
     517 Gly...GlyThrThrGlyThrAsnThrIleSer................. 526

     837 AAAGTGGTTAATTAATGGGGTATTGCAAACAGGCAACCCCTATATA...G 883
               :::   :::          :::::::::||||||      |||   :
     527 .......ValThrLysAspGlyIleLysAlaGlyAsnLysAlaIleThrA 541

     884 GAAAAAGCAATGGCTTCCAGCTAGTTCGTAAAGATTGGTTCTATGATGAA 933
         ::   ::::::|||::::::                          ||||||:::
     541 snValAlaSerGlyLeuArgAla.................TyrAspAsp 551

     934 ATCTTTGCTGGAGATACCCATTCAGTATTCTACGAACCACATCAAAATGG 983

     552 Ala.............................................. 552

     984 GAAATACTTTTTTAACGACAATAATAATGGCGCAGGAAAAATCGATGCCA 1033
                   |||::::      ||||||:::|||      :::
     553 .......AsnPheAspValLeuAsnAsnSerAlaThrAspLeu....... 564
```

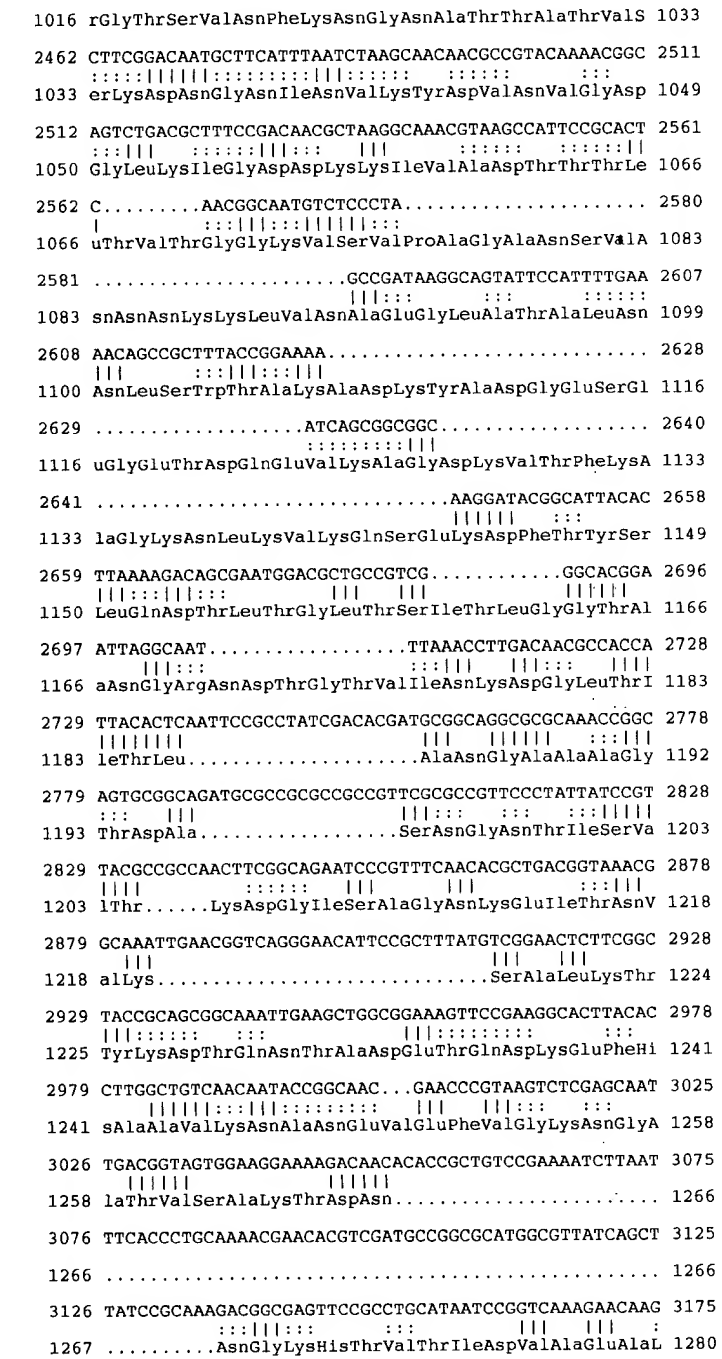

1034 AACATAAACACTATTCTCTACCTTATAGATTAAAAACACGAACCGTTCAA 1083
565 ..AsnArgHisValGluAspAlaTyrLys.....Gly 574
1084 TTGTTTAATGTTTCTTTATCCGAGACAGCAAGAGAACCTGTTTAT..... 1128
575 LeuLeuAsnLeuAsnGluLysAsnAlaAsnLysGlnProLeuValThrAs 591
1129 .CATGCTGCAGGTGGGGTCAACAGTTATCGACCCAGACTGAATAATGGAG 1177
591 pSerThrAlaAlaThrValGlyAspLeuArgLys..... 602
1178 AAAATATTTCCTTTATTGACAAAGGAAAAGGTGAATTGATACTTACCAGC 1227
603 .....LeuGlyTrpValValSerThrLys 610
1228 AACATCAACCAAGGCGCGGGCGGTTTGTATTTTGAGGGTAATTTTACGGT 1277
611 Asn.....GlyThrLysGluGluSerAsn...GlnVa 620
1278 CTCGCCTAAAAACAACGAAACGTGGCAAGGCGCGGGCGTTCATATCAGTG 1327
620 lLysGlnAlaAspGluValLeuPheThrGlyAlaGly..... 632
1328 ATGGCAGTACCGTTACTTGGAAAGTAAACGGCGTGGCAAACGACCGCCTG 1377
633 ..AlaAlaThrValThrSerLys..... 639
1378 TCCAAAATCGGCAAAGGCACGCTGCTGGTTCAAGCCAAAGGGGAAAACCA 1427
640 SerGluAsnGlyLysHisThrIle..... 647
1428 AGGCTCGGTCAGCGTGGGCGACGGTAAAGTCATCTTAGATCAGCAGGCGG 1477
648 ....ThrValSerValAlaGluThrLys..... 655
1478 ACGATCAAGGCAAAAAACAAGCCTTTAGTGAAATCGGCTTGGTCAGCGGC 1527
656 .....AlaAspCysGlyLeuGluLysAsp 663
1528 AGGGGGACGGTGCAACTGAATGCCGATAATCAGTTCAACCCCGACAAACT 1577
664 GlyAspThrIleLysLeuLysValAspAsnGlnAsnThrAspAsnValLe 680
1578 CTATTTCGGCTTTCGCGGCGGACGTTTGGATTTGAACGGGCATTCGCTTT 1627
680 uThrValGlyAsnAsnGlyThrAlaValThrLysGlyGlyPheGluThrV 697
1628 CGTTCCACCGCATTCAAAATACCGATGAAGGGGCGATGATTGTCAAC....1674
697 alLysThrGlyAlaThrAspAlaAspArgGlyLysValThrValLysAsp 713
1675 .....CACAATCAAGACAAAGAA..... 1692
714 AlaThrAlaAsnAspAlaAspLysLysValAlaThrValLysAspValAl 730
1693 .....TCCACCGTTACCATTACAGGCAATAAAGATATTACTA 1729
730 aThrAlaIleAsnSerAlaAlaThrPheValLysThrGluAsnLeuThrT 747
1730 CAACC...GGCAATAACAACAACTTGGATAGCAAAAAAGAAATTGCCTAC 1776
747 hrSerIleAspGluAspAsnProThrAspAsnGlyLysAspAspAlaLeu 763
1777 AACGGTTGG.....TTTGGCGAGAAAGATGCAACCAA 1808
764 LysAlaGlyAspThrLeuThrPheLysAlaGlyLysAsnLeuLysValLy 780
1809 AACGAACGGGCGG.....CTCAATCTGAATTACC 1837
780 sArgAspGlyLysAsnIleThrPheAspLeuAlaLysAsnLeuGluValL 797
1838 AACCGGAAGAAGCGGATCGCACTTTACTGCTTTCCGGCGGAACAAATTTA 1887
797 ysThrAlaLysValSerAspThrLeuThrIleGlyGlyAsnThrProThr 813
1888 AACGGC.....AATATCACGCAAACAAACGGCAA 1916
814 GlyGlyThrThrAlaThrProLysValAsnIleThrSerThrAlaAspGl 830
1917 ACTGTTTTTCAGCGGCAGACCGACACCGCACGCCTACAATCATTTAGGAA 1966
830 yLeuAsnPheAla...LysGluThrAlaAspAla.....SerGlyS 843
1967 GCGGGTGGTCAAAAATGGAAGGTATCCCACAAGGAGAAATCGTGTGGGAC 2016
843 erLysAsnValTyrLeuLysGlyIleAla..... 852
2017 AACGATTGGATCGACCGCACATTTAAAGCGGAAAACTTCCATATTCAG.. 2064
853 ThrThrLeuThrGluProSerAlaGlyAlaLysSerSerHisValAspLe 869
2065 .GGCGGACAAGCGGTGGTTTCCCGCAATGTTGCCAAAGTGGAA..... 2106
869 uAsnValAspAlaThrLysLysSerAsnAlaAlaSerIleGluAspValL 886
2107 .....GGCGATTGGCATTTAAGCAATCACGCCCAAGCAGTTTCGGTGTC 2151
886 euArgAlaGlyTrpAsnIleGlnGlyAsnGlyAsnAsnValAspTyrVal 902
2152 GCACCGCATCAAAGCCACACAATCTGTACACGTTCGGACTGGACGGGTCT 2201
903 AlaThrTyrAsp.....ThrValAsnPheThrAspAspSerThrGlyTh 917
2202 GACAAGTTGTACCGAAAAAACCATTACCGACGATAAA.....GTGATTG 2245
917 rThrThrValThrValThrGlnLysAlaAspGlyLysGlyAlaAspValL 934
2246 CTTCATTGAGCAAGACCGACATC..... 2268
934 ysIleGlyAlaLysThrSerValIleLysAspHisAsnGlyLysLeuPhe 950
2269 .....AGAGGCAATGTCAGCCTTGCCGA 2291
951 ThrGlyLysAspLeuLysAspAlaAsnAsnGlyAlaThrValSerGluAs 967
2292 TCACGCTCATTTAAATCTCACAGGACTTGCCACACTCAACGGCAATCTTA 2341
967 pAspGlyLysAspThrGlyThrGlyLeuValThrAlaLysThrValIleA 984
2342 GTGCAGGCGGAGACACGCACTATACGGTTACG...CGCAACGCCACCCAA 2388
984 spAlaValAsnLysSerGlyTrpArgValThrGlyGluGlyAlaThrAla 1000
2389 AACGGCAACCTCAGCCTCGTG.....GGCAATGCCCAAGCAACATTTAA 2432
1001 GluThrGlyAlaThrAlaValAsnAlaGlyAsnAlaGluThrValThrSe 1017
2433 TCAAGCCACATTA.....AACGGCAAC.....ACATCGG 2461
1017 rGlyThrSerValAsnPheLysAsnGlyAsnAlaThrThrAlaThrValS 1034
2462 CTTCGGACAATGCTTCATTTAATCTAAGCAACAACGCCGTACAAAACGGC 2511
1034 erLysAspAsnGlyAsnIleAsnValLysTyrAspValAsnValGlyAsp 1050
2512 AGTCTGACGCTTTCCGACAACGCTAAGGCAAACGTAAGCCATTCCGCACT 2561
1051 GlyLeuLysIleGlyAspAspLysLysIleValAlaAspThrThrThrLe 1067
2562 C.....AACGGCAATGTCTCCCTA..... 2580
1067 uThrValThrGlyGlyLysValSerValProAlaGlyAlaAsnSerValA 1084
2581 .....GCCGATAAGGCAGTATTCCATTTTGAA 2607



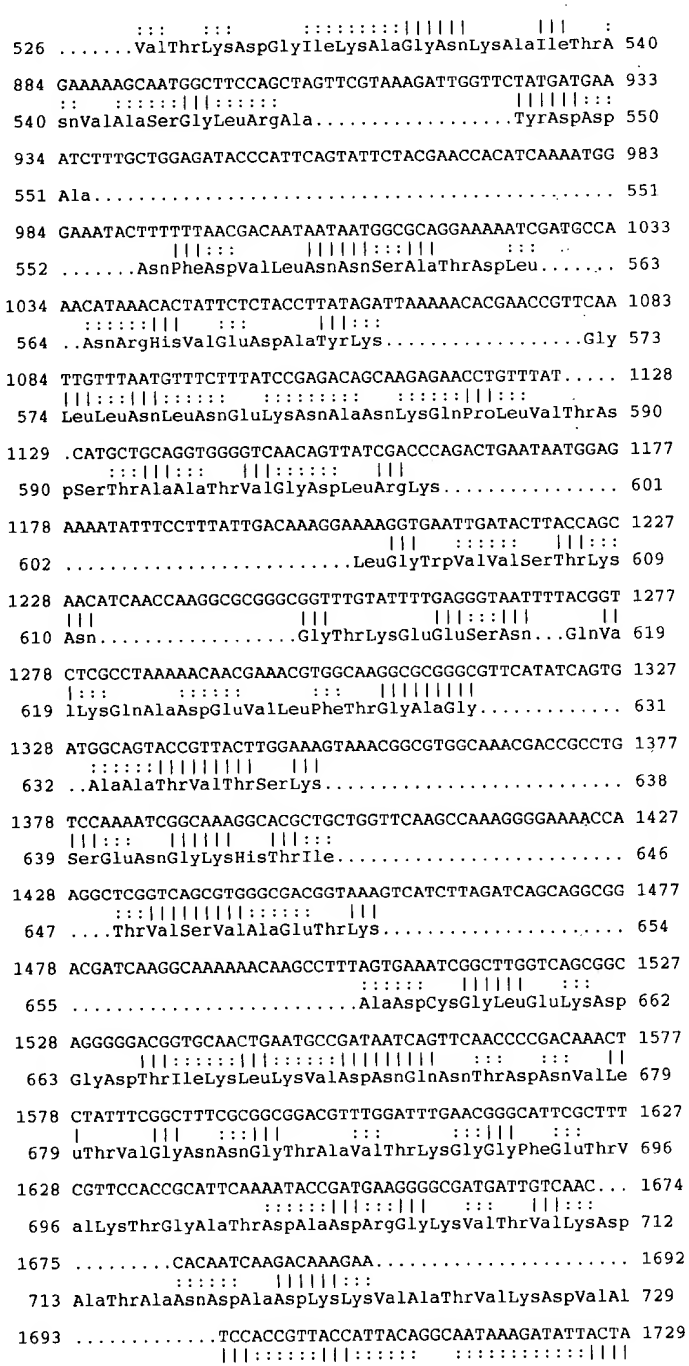
```
   4213 GCCGCTTCCGGCAAAGTCCGAACGCGCGTCAATACCGCCGTATTGGCGCA 4262
        :::||||||         |||   :::::::::::|||   :::::
   1549 PheValSerGlyAspLysAspThr...ThrSerValThrValGluSerLy 1564

   4263 GGATTTCGGCAAAACCCGCAGTGCGGAATGGGGCGTAAACGCCGAAATC. 4311
        :|||   ||||||      :::::::::   |||:::::::::::::::
   1564 sAspAsnGlyLysArgThrGluValLysIleGlyAlaLysThrSerValI 1581

   4312 ................AAAGGTTTCACGCTGTCCCTCCACGCTGCCGCC 4344
                        |||   ||||||
   1581 leLysAspHisAsnGlyLysLeuPheThr..................... 1590

   4345 GCCAAGGGGCCGCAATTGGAAGCGCAGCACAGCGCGGGCATC 4386
            |||   :::|||:::      ::::::   |||:::
   1591 ......GlyLysGluLeuLysAspAlaAsnAsnAsnGlyVal 1602

seq_name: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:US-08-409-995-4

seq_documentation_block:
; Sequence 4, Application US/08409995
; Patent No. 5646259
;  GENERAL INFORMATION:
;    APPLICANT:  Barenkamp, Stephen I.
;    APPLICANT:  St. Geme III, Joseph W.
;    TITLE OF INVENTION:  Haemophilus Adhesion Proteins
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Flehr, Hohbach, Test, Albritton & Herbert
;      STREET:  Four Embarcadero Center, Suite 3400
;      CITY:  San Francisco
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94111-4187
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/409,995
;      FILING DATE:  24-MAR-1995
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Silva, Robin M.
;      REGISTRATION NUMBER:  38,304
;      REFERENCE/DOCKET NUMBER:  A-61053/RFT
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 781-1989
;      TELEFAX:  (415) 398-3249
;      TELEX:  910 277299
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1912 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  unknown
US-08-409-995-4


alignment_scores:
            Quality:  248.50            Length:   1577
              Ratio:  0.350               Gaps:     70
 Percent Similarity:  44.959   Percent Identity: 19.784

alignment_block:
 US-09-303-518D-653 x US-08-409-995-4   ..

 Align seg 1/1  to: US-08-409-995-4  from: 1  to: 1912

       7 ACAACCGACAAACGGACAACCGAA.....ACACACCGCAAAGCCCC 47
         |||   ||||||:::|||   :::         |||   ::::::
     245 ThrGlyAspLysAsnThrLeuAspValValLeuThrAlaLysGluAsn** 261

  48 TAAAACCGGCCGCATCCGCTTCTCGCCCGCTTACTTAGCCATATGCCTGT 97
     ||||||   :::::::::|||:::|||         :::|||      :
 261 *LysThrThrGluValLysPheThrProLysThrSerValIleLysGluL 278

  98 CGTTCGGCATTCTGCCCCAAGCCCGGGCGGGACACACTTATTTCGGCATC 147
     ::   |||   |||      ::::::   ::::::          :::
 278 ysAspGlyLysLeuPheThrGlyLysGluAsnAsnAspThrAsnLysVal 294

 148 AACTACCAATACTATCGCGACTTTGCCGAAAATAAAGGCAAGTTTGCAGT 197
     :::   :::         |||   :::::::::   :::   ::::::::
 295 ThrSerAsnThrAlaThrAspAsnThrAspGluGlyAsnGlyLeuValTh 311

 198 CGGGGCGAAAGATATTGAGGTTTACAACAAAAAAGGGGAGTTGGTCGGCA 247
     ::::         |||:::::::   ||||||   |||      |||
 311 rAlaLysAlaValIleAspAlaValAsnLysAlaGlyTrpArgValLysT 328

 248 AATCGATGACGAAAGCCCCGATGATTGATTTT...TCTGTGGTATCGCGT 294
       :::   :::::::::         ||||||   :::|||:::|||
 328 hrThrThrAlaAsnGlyGlnAsnGlyAspPheAlaThrValAlaSerGly 344

 295 AACGGCGTGGCGGCATTGGCGGGCGATCAATATATTGTGAGCGTGGCACA 344
     ::::::|||:::      :::|||||||         :::|||||||:::
 345 ThrAsnValThrPheGluSerGlyAspGlyThrThrAlaSerValThrLy 361

 345 TAACGGCGGCTATAACAATGTTGATTTTGGTGCGGAGGGAAGCAATCCCG 394
       :::   :::   |||:::
 361 sAspThrAsnGlyAsnGly.............................. 367

 395 ATCAGCACCGCTTTTCTTACCAAATTGTGAAAAGAAATAATTATAAAGCA 444
                           |||:::   :::   :::   |||:::
 368 .......................IleThrValLysTyrAspAlaLysVal 376

 445 GGGACTAACGGCCATCCTTATGGCGGCGATTATCATATGCCGCGTTTGCA 494
     |||   :::|||      :::   :::|||
 377 Gly...AspGlyLeuLysPheAspSerAsp..................Ly 386

 495 CAAATTTGTCACAGATGCAGAACCTGTTGAGATGACCAGTTATATGGATG 544
      |||:::|||:::|||:::      :::   ::::||           |
 386 sLysIleValAlaAspThrThrAlaLeuThrValThr............G 399

 545 GGTGGAAATACGCTGATTTAAATAAATACCCTGATCGTGTTCGAATC... 591
     ||   |||   |||::::::   |||      |||:::   ::::::
 399 lyGlyLysValAlaGluIleAlaLysGluAspAspLysLysLysLeuVal 415

 592 .....................GGAGCAGGCAGACAATATTGGCGG..... 615
                           :::   |||:::      |||:::
 416 AsnAlaGlyAspLeuValThrAlaLeuGlyAsnLeuSerTrpLysAlaLy 432

 616 .......TCTGATGAAGAC.........GAACCCAATAACCGCGAAAGTT 649
           :::|||   |||         |||      ::::::::::::
 432 sAlaGluAlaAspThrAspGlyAlaLeuGluGlyIleSerLysAspGlnG 449

 650 CATATCATATTGCAAGCGCATATTCTTGGCTCGTCGGTGGCAAT...ACC 696
     ::         :::::::::   ::::::   :::|||   |||
 449 luValLysAlaGlyGluThrValThrPheLysAlaGlyLysAsnLeuLys 465

 697 TTTGCACAAAATGGATCAGGT........................... 717
           |||:::|||::::::
 466 ValLysGlnAspGlyAlaAsnPheThrTyrSerLeuGlnAspAlaLeuTh 482

 718 .....................GGTGGCACAGTCAACTTAGGTAGCGAAA 745
                           |||||||||:::|||   |||::::::
 482 rGlyLeuThrSerIleThrLeuGlyGlyThrThrAsnGlyGlyAsnAspA 499

 746 AA.........ATTAAACATAGCCCATATGGTTTTTTACCAACAGGAGGC 786
               |||:::   :::         :::   |||:::|||:::
 499 laLysThrValIleAsnLysAspGlyLeuThrIleThrProAlaGlyAsn 515

 787 TCATTTGGCGACAGTGGCTCACCAATGTTTATCTATGATGCCCAAAAGCA 836
     :::   |||   :::|||:::      :::
 516 Gly...GlyThrThrGlyThrAsnThrIleSer................. 525
```

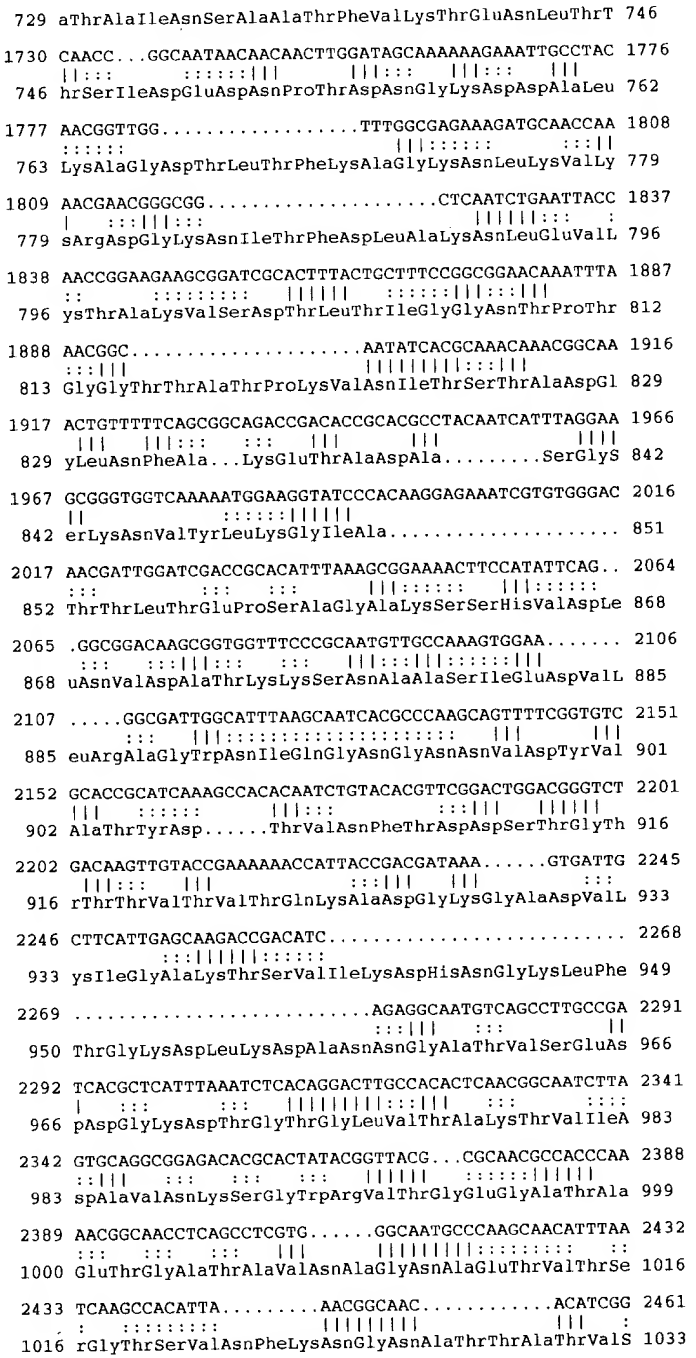
```
 837 AAAGTGGTTAATTAATGGGGTATTGCAAACAGGCAACCCCTATATA...G 883
            :::  :::      :::::::::|||||      |||   :
 526 .......ValThrLysAspGlyIleLysAlaGlyAsnLysAlaIleThrA 540

 884 GAAAAAGCAATGGCTTCCAGCTAGTTCGTAAAGATTGGTTCTATGATGAA 933
      ::   ::::::|||:::::                    |||||||:::
 540 snValAlaSerGlyLeuArgAla..................TyrAspAsp 550

 934 ATCTTTGCTGGAGATACCCATTCAGTATTCTACGAACCACATCAAAATGG 983

 551 Ala............................................... 551

 984 GAAATACTTTTTTAACGACAATAATAATGGCGCAGGAAAAATCGATGCCA 1033
                ||||:::    |||||||:::|||      :::
 552 .......AsnPheAspValLeuAsnAsnSerAlaThrAspLeu....... 563

1034 AACATAAACACTATTCTCTACCTTATAGATTAAAAACACGAACCGTTCAA 1083
        ::::::|||    :::       |||:::
 564 ..AsnArgHisValGluAspAlaTyrLys...................Gly 573

1084 TTGTTTAATGTTTCTTTATCCGAGACAGCAAGAGAACCTGTTTAT..... 1128
     |||:::|||::::::    :::::::::    ::::::|||:::
 574 LeuLeuAsnLeuAsnGluLysAsnAlaAsnLysGlnProLeuValThrAs 590

1129 .CATGCTGCAGGTGGGGTCAACAGTTATCGACCCAGACTGAATAATGGAG 1177
         :::|||:::    |||::::::    |||
 590 pSerThrAlaAlaThrValGlyAspLeuArgLys................ 601

1178 AAAATATTTCCTTTATTGACAAAGGAAAAGGTGAATTGATACTTACCAGC 1227
                                 |||    ::::::    |||:::
 602 ..........................LeuGlyTrpValValSerThrLys 609

1228 AACATCAACCAAGGCGCGGGCGGTTTGTATTTGAGGGTAATTTTACGGT 1277
      |||                  |||        |||:::|||       ||
 610 Asn..................GlyThrLysGluGluSerAsn...GlnVa 619

1278 CTCGCCTAAAAACAACGAAACGTGGCAAGGCGCGGGCGTTCATATCAGTG 1327
      |:::      ::::::       :::    |||||||||
 619 lLysGlnAlaAspGluValLeuPheThrGlyAlaGly............. 631

1328 ATGGCAGTACCGTTACTTGGAAAGTAAACGGCGTGGCAAACGACCGCCTG 1377
        ::::::||||||||   |||
 632 ..AlaAlaThrValThrSerLys........................... 638

1378 TCCAAAATCGGCAAAGGCACGCTGCTGGTTCAAGCCAAAGGGGAAAACCA 1427
     |||:::    ||||||    |||:::
 639 SerGluAsnGlyLysHisThrIle.......................... 646

1428 AGGCTCGGTCAGCGTGGGCGACGGTAAAGTCATCTTAGATCAGCAGGCGG 1477
        :::||||||||||::::::   |||
 647 ....ThrValSerValAlaGluThrLys...................... 654

1478 ACGATCAAGGCAAAAAACAAGCCTTTAGTGAAATCGGCTTGGTCAGCGGC 1527
                               ::::::    ||||||   :::
 655 .........................AlaAspCysGlyLeuGluLysAsp 662

1528 AGGGGGACGGTGCAACTGAATGCCGATAATCAGTTCAACCCCGACAAACT 1577
           |||::::::|||::::::|||||||||    :::    :::   ||
 663 GlyAspThrIleLysLeuLysValAspAsnGlnAsnThrAspAsnValLe 679

1578 CTATTTCGGCTTTCGCGGCGGACGTTTGGATTTGAACGGGCATTCGCTTT 1627
      |       |||    :::|||       :::    :::|||    :::
 679 uThrValGlyAsnAsnGlyThrAlaValThrLysGlyGlyPheGluThrV 696

1628 CGTTCCACCGCATTCAAAATACCGATGAAGGGGCGATGATTGTCAAC... 1674
                         ::::::|||:::|||    :::    |||:::
 696 alLysThrGlyAlaThrAspAlaAspArgGlyLysValThrValLysAsp 712

1675 .........CACAATCAAGACAAAGAA....................... 1692
              ::::::    |||||||:::
 713 AlaThrAlaAsnAspAlaAspLysLysValAlaThrValLysAspValAl 729

1693 .............TCCACCGTTACCATTACAGGCAATAAAGATATTACTA 1729
```

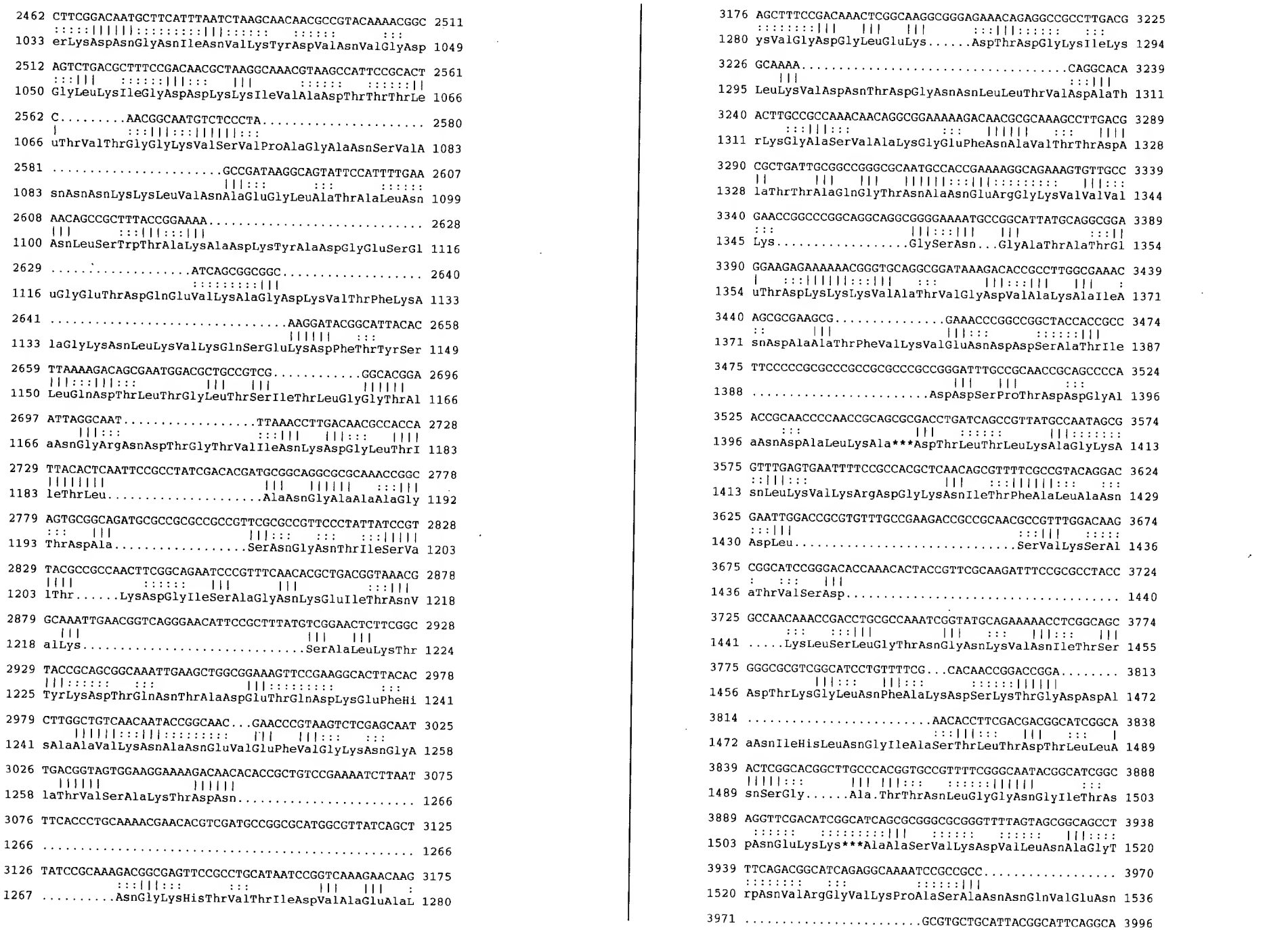
```
          |||::::::|||::::::    ::::::::::::||||
 729 aThrAlaIleAsnSerAlaAlaThrPheValLysThrGluAsnLeuThrT 746

1730 CAACC...GGCAATAACAACAACTTGGATAGCAAAAAAGAAATTGCCTAC 1776
      |||:::       ::::::|||       |||:::   |||:::    |||
 746 hrSerIleAspGluAspAsnProThrAspAsnGlyLysAspAspAlaLeu 762

1777 AACGGTTGG.................TTTGGCGAGAAAGATGCAACCAA 1808
      ::::::                        |||::::::      :::||
 763 LysAlaGlyAspThrLeuThrPheLysAlaGlyLysAsnLeuLysValLy 779

1809 AACGAACGGGCGG.....................CTCAATCTGAATTACC 1837
      |    :::|||:::                        ||||||:::   :
 779 sArgAspGlyLysAsnIleThrPheAspLeuAlaLysAsnLeuGluValL 796

1838 AACCGGAAGAAGCGGATCGCACTTTACTGCTTTCCGGCGGAACAAATTTA 1887
      ::        :::::::::      ||||||    ::::::|||:::|||
 796 ysThrAlaLysValSerAspThrLeuThrIleGlyGlyAsnThrProThr 812

1888 AACGGC......................AATATCACGCAAACAAACGGCAA 1916
      :::|||                    |||||||||:::|||
 813 GlyGlyThrThrAlaThrProLysValAsnIleThrSerThrAlaAspGl 829

1917 ACTGTTTTTCAGCGGCAGACCGACACCGCACGCCTACAATCATTTAGGAA 1966
       |||    |||:::    :::    |||       |||          ||||
 829 yLeuAsnPheAla...LysGluThrAlaAspAla.........SerGlyS 842

1967 GCGGGTGGTCAAAAATGGAAGGTATCCCACAAGGAGAAATCGTGTGGGAC 2016
      ||            ::::::||||||
 842 erLysAsnValTyrLeuLysGlyIleAla..................... 851

2017 AACGATTGGATCGACCGCACATTTAAAGCGGAAAACTTCCATATTCAG.. 2064
      :::         :::    :::       |||::::::    |||::::::
 852 ThrThrLeuThrGluProSerAlaGlyAlaLysSerSerHisValAspLe 868

2065 .GGCGGACAAGCGGTGGTTTCCCGCAATGTTGCCAAAGTGGAA....... 2106
      :::    :::|||:::    :::    |||:::|||::::::|||
 868 uAsnValAspAlaThrLysLysSerAsnAlaAlaSerIleGluAspValL 885

2107 .....GGCGATTGGCATTTAAGCAATCACGCCCAAGCAGTTTTCGGTGTC 2151
          :::    |||:::::::::::::::::::::    |||      |||
 885 euArgAlaGlyTrpAsnIleGlnGlyAsnGlyAsnAsnValAspTyrVal 901

2152 GCACCGCATCAAAGCCACACAATCTGTACACGTTCGGACTGGACGGGTCT 2201
      |||    ::::::       |||:::          :::|||    ||||||
 902 AlaThrTyrAsp......ThrValAsnPheThrAspAspSerThrGlyTh 916

2202 GACAAGTTGTACCGAAAAAACCATTACCGACGATAAA......GTGATTG 2245
       |||:::    |||            :::|||    |||         :::
 916 rThrThrValThrValThrGlnLysAlaAspGlyLysGlyAlaAspValL 933

2246 CTTCATTGAGCAAGACCGACATC........................... 2268
              :::||||||::::::
 933 ysIleGlyAlaLysThrSerValIleLysAspHisAsnGlyLysLeuPhe 949

2269 ..........................AGAGGCAATGTCAGCCTTGCCGA 2291
                                :::|||    :::         ||
 950 ThrGlyLysAspLeuLysAspAlaAsnAsnGlyAlaThrValSerGluAs 966

2292 TCACGCTCATTTAAATCTCACAGGACTTGCCACACTCAACGGCAATCTTA 2341
      |    :::         :::    |||||||||:::|||    :::       ::::
 966 pAspGlyLysAspThrGlyThrGlyLeuValThrAlaLysThrValIleA 983

2342 GTGCAGGCGGAGACACGCACTATACGGTTACG...CGCAACGCCACCCAA 2388
      ::|||    :::    :::    :::    ||||||    ::::::||||||
 983 spAlaValAsnLysSerGlyTrpArgValThrGlyGluGlyAlaThrAla 999

2389 AACGGCAACCTCAGCCTCGTG......GGCAATGCCCAAGCAACATTTAA 2432
      :::    :::    :::    |||        |||||||||:::::::::    ::
1000 GluThrGlyAlaThrAlaValAsnAlaGlyAsnAlaGluThrValThrSe 1016

2433 TCAAGCCACATTA.........AACGGCAAC...........ACATCGG 2461
      :    :::::::::               |||||||||           |||    :
```


1016 rGlyThrSerValAsnPheLysAsnGlyAsnAlaThrThrAlaThrValS 1033
2462 CTTCGGACAATGCTTCATTTAATCTAAGCAACAACGCCGTACAAAACGGC 2511
1033 erLysAspAsnGlyAsnIleAsnValLysTyrAspValAsnValGlyAsp 1049
2512 AGTCTGACGCTTTCCGACAACGCTAAGGCAAACGTAAGCCATTCCGCACT 2561
1050 GlyLeuLysIleGlyAspAspLysLysIleValAlaAspThrThrThrLe 1066
2562 C.....AACGGCAATGTCTCCCTA..... 2580
1066 uThrValThrGlyGlyLysValSerValProAlaGlyAlaAsnSerValA 1083
2581 .....GCCGATAAGGCAGTATTCCATTTTGAA 2607
1083 snAsnAsnLysLysLeuValAsnAlaGluGlyLeuAlaThrAlaLeuAsn 1099
2608 AACAGCCGCTTTACCGGAAAA..... 2628
1100 AsnLeuSerTrpThrAlaLysAlaAspLysTyrAlaAspGlyGluSerGl 1116
2629 .....ATCAGCGGCGGC..... 2640
1116 uGlyGluThrAspGlnGluValLysAlaGlyAspLysValThrPheLysA 1133
2641 .....AAGGATACGGCATTACAC 2658
1133 laGlyLysAsnLeuLysValLysGlnSerGluLysAspPheThrTyrSer 1149
2659 TTAAAAGACAGCGAATGGACGCTGCCGTCG.....GGCACGGA 2696
1150 LeuGlnAspThrLeuThrGlyLeuThrSerIleThrLeuGlyGlyThrAl 1166
2697 ATTAGGCAAT.....TTAAACCTTGACAACGCCACCA 2728
1166 aAsnGlyArgAsnAspThrGlyThrValIleAsnLysAspGlyLeuThrI 1183
2729 TTACACTCAATTCCGCCTATCGACACGATGCGGCAGGCGCGCAAACCGGC 2778
1183 leThrLeu.....AlaAsnGlyAlaAlaAlaGly 1192
2779 AGTGCGGCAGATGCGCCGCGCCGCCGTTCGCGCCGTTCCCTATTATCCGT 2828
1193 ThrAspAla.....SerAsnGlyAsnThrIleSerVa 1203
2829 TACGCCGCCAACTTCGGCAGAATCCCGTTTCAACACGCTGACGGTAAACG 2878
1203 lThr.....LysAspGlyIleSerAlaGlyAsnLysGluIleThrAsnV 1218
2879 GCAAATTGAACGGTCAGGGAACATTCCGCTTTATGTCGGAACTCTTCGGC 2928
1218 alLys.....SerAlaLeuLysThr 1224
2929 TACCGCAGCGGCAAATTGAAGCTGGCGGAAAGTTCCGAAGGCACTTACAC 2978
1225 TyrLysAspThrGlnAsnThrAlaAspGluThrGlnAspLysGluPheHi 1241
2979 CTTGGCTGTCAACAATACCGGCAAC...GAACCCGTAAGTCTCGAGCAAT 3025
1241 sAlaAlaValLysAsnAlaAsnGluValGluPheValGlyLysAsnGlyA 1258
3026 TGACGGTAGTGGAAGGAAAAGACAACACACCGCTGTCCGAAAATCTTAAT 3075
1258 laThrValSerAlaLysThrAspAsn..... 1266
3076 TTCACCCTGCAAAACGAACACGTCGATGCCGGCGCATGGCGTTATCAGCT 3125
1266 1266
3126 TATCCGCAAAGACGGCGAGTTCCGCCTGCATAATCCGGTCAAAGAACAAG 3175
1267 .....AsnGlyLysHisThrValThrIleAspValAlaGluAlaL 1280
3176 AGCTTTCCGACAAACTCGGCAAGGCGGGAGAAACAGAGGCCGCCTTGACG 3225
1280 ysValGlyAspGlyLeuGluLys.....AspThrAspGlyLysIleLys 1294
3226 GCAAAA.....CAGGCACA 3239
1295 LeuLysValAspAsnThrAspGlyAsnAsnLeuLeuThrValAspAlaTh 1311
3240 ACTTGCCGCCAAACAACAGGCGGAAAAAGACAACGCGCAAAGCCTTGACG 3289
1311 rLysGlyAlaSerValAlaLysGlyGluPheAsnAlaValThrThrAspA 1328
3290 CGCTGATTGCGGCCGGGCGCAATGCCACCGAAAAGGCAGAAAGTGTTGCC 3339
1328 laThrThrAlaGlnGlyThrAsnAlaAsnGluArgGlyLysValValVal 1344
3340 GAACCGGCCCGGCAGGCAGGCGGGGAAAATGCCGGCATTATGCAGGCGGA 3389
1345 Lys.....GlySerAsn...GlyAlaThrAlaThrGl 1354
3390 GGAAGAGAAAAAACGGGTGCAGGCGGATAAAGACACCGCCTTGGCGAAAC 3439
1354 uThrAspLysLysLysValAlaThrValGlyAspValAlaLysAlaIleA 1371
3440 AGCGCGAAGCG.....GAAACCCGGCCGGCTACCACCGCC 3474
1371 snAspAlaAlaThrPheValLysValGluAsnAspAspSerAlaThrIle 1387
3475 TTCCCCCGCGCCCGCCGCGCCCGCCGGGATTTGCCGCAACCGCAGCCCCA 3524
1388 .....AspAspSerProThrAspAspGlyAl 1396
3525 ACCGCAACCCCAACCGCAGCGCGACCTGATCAGCCGTTATGCCAATAGCG 3574
1396 aAsnAspAlaLeuLysAla***AspThrLeuThrLeuLysAlaGlyLysA 1413
3575 GTTTGAGTGAATTTTCCGCCACGCTCAACAGCGTTTTCGCCGTACAGGAC 3624
1413 snLeuLysValLysArgAspGlyLysAsnIleThrPheAlaLeuAlaAsn 1429
3625 GAATTGGACCGCGTGTTTGCCGAAGACCGCCGCAACGCCGTTTGGACAAG 3674
1430 AspLeu.....SerValLysSerAl 1436
3675 CGGCATCCGGGACACCAAACACTACCGTTCGCAAGATTTCCGCGCCTACC 3724
1436 aThrValSerAsp..... 1440
3725 GCCAACAAACCGACCTGCGCCAAATCGGTATGCAGAAAAACCTCGGCAGC 3774
1441 .....LysLeuSerLeuGlyThrAsnGlyAsnLysValAsnIleThrSer 1455
3775 GGGCGCGTCGGCATCCTGTTTTCG...CACAACCGGACCGGA..... 3813
1456 AspThrLysGlyLeuAsnPheAlaLysAspSerLysThrGlyAspAspAl 1472
3814 .....AACACCTTCGACGACGGCATCGGCA 3838
1472 aAsnIleHisLeuAsnGlyIleAlaSerThrLeuThrAspThrLeuLeuA 1489
3839 ACTCGGCACGGCTTGCCCACGGTGCCGTTTTCGGGCAATACGGCATCGGC 3888
1489 snSerGly.....Ala.ThrThrAsnLeuGlyGlyAsnGlyIleThrAs 1503
3889 AGGTTCGACATCGGCATCAGCGCGGGCGCGGGTTTTAGTAGCGGCAGCCT 3938
1503 pAsnGluLysLys***AlaAlaSerValLysAspValLeuAsnAlaGlyT 1520
3939 TTCAGACGGCATCAGAGGCAAAATCCGCCGCC..... 3970
1520 rpAsnValArgGlyValLysProAlaSerAlaAsnAsnGlnValGluAsn 1536

```
     3971 .....................GCGTGCTGCATTACGGCATTCAGGCA 3996
                               :::    :::::::::      ::
     1537 IleAspPheValAlaThrTyrAspThrValAspPheValSerGlyAspLy 1553

     3997 AGATACCGCGCAGGTTTCGGCGGATTCGGCATCGAACCGCACATCGGCGC 4046
          :||||||::::::|||:::::::::|||   :::|||         ::::
     1553 sAspThrThrSerValThrValGluSerLysAspAsnGlyLysArgThrG 1570

     4047 AACGCGCTATTTCGTCCAAAAAGCGGATT 4075
          ::      |||:::::::|||      :::
     1570 luValLysIleGlyAlaLysThrSerVal 1579

seq_name: /cgn2_6/ptodata/1/iaa/6A_COMB.pep:US-08-685-467-4

seq_documentation_block:
; Sequence 4, Application US/08685467
; Patent No. 6060059
;  GENERAL INFORMATION:
;    APPLICANT:  St. Geme III, Joseph W.
;    APPLICANT:  Barenkamp, Stephen J.
;    TITLE OF INVENTION:  HAEMOPHILUS ADHESION PROTEINS
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Flehr, Hohbach, Test, Albritton & Herbert
;      STREET:  Four Embarcadero Center, Suite 3400
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  United States
;      ZIP:  94111-4187
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/685,467
;      FILING DATE:  22-JUL-1996
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/409,995
;      FILING DATE:  24-MAR-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Silva, Robin M.
;      REGISTRATION NUMBER:  38,304
;      REFERENCE/DOCKET NUMBER:  A-61053-2/RFT/RMS
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 781-1989
;      TELEFAX:  (415) 398-3249
;      TELEX:  910 277299
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1912 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  protein
US-08-685-467-4


alignment_scores:
            Quality:  248.50              Length:   1577
              Ratio:  0.350                 Gaps:     70
 Percent Similarity:  44.959   Percent Identity: 19.784

alignment_block:
 US-09-303-518D-653 x US-08-685-467-4   ..

 Align seg 1/1  to: US-08-685-467-4  from: 1  to: 1912

       7 ACAACCGACAAACGGACAACCGAA.........ACACACCGCAAAGCCCC 47
         |||   ||||||:::|||   :::         |||   ::::::
     245 ThrGlyAspLysAsnThrLeuAspValValLeuThrAlaLysGluAsn** 261

      48 TAAAACCGGCCGCATCCGCTTCTCGCCCGCTTACTTAGCCATATGCCTGT 97
         ||||||   :::::::::|||:::|||         :::|||      :
     261 *LysThrThrGluValLysPheThrProLysThrSerValIleLysGluL 278

      98 CGTTCGGCATTCTGCCCCAAGCCCGGGCGGGACACACTTATTTCGGCATC 147
         ::   |||   |||       ::::::   ::::::            :::
     278 ysAspGlyLysLeuPheThrGlyLysGluAsnAsnAspThrAsnLysVal 294

     148 AACTACCAATACTATCGCGACTTTGCCGAAAATAAAGGCAAGTTTGCAGT 197
         :::   :::         |||   :::::::::   :::   ::::::::
     295 ThrSerAsnThrAlaThrAspAsnThrAspGluGlyAsnGlyLeuValTh 311

     198 CGGGGCGAAAGATATTGAGGTTTACAACAAAAAAGGGGAGTTGGTCGGCA 247
         ::::         |||:::::::   ||||||   |||      |||
     311 rAlaLysAlaValIleAspAlaValAsnLysAlaGlyTrpArgValLysT 328

     248 AATCGATGACGAAAGCCCCGATGATTGATTTT...TCTGTGGTATCGCGT 294
         :::   :::::::::         ||||||   :::|||:::|||
     328 hrThrThrAlaAsnGlyGlnAsnGlyAspPheAlaThrValAlaSerGly 344

     295 AACGGCGTGGCGGCATTGGCGGGCGATCAATATATTGTGAGCGTGGCACA 344
         ::::::|||:::      :::||||||         :::|||||:::
     345 ThrAsnValThrPheGluSerGlyAspGlyThrThrAlaSerValThrLy 361

     345 TAACGGCGGCTATAACAATGTTGATTTTGGTGCGGAGGGAAGCAATCCCG 394
         :::   :::   |||:::
     361 sAspThrAsnGlyAsnGly..............................  367

     395 ATCAGCACCGCTTTTCTTACCAAATTGTGAAAAGAAATAATTATAAAGCA 444
                               |||:::   :::   :::   |||:::
     368 .....................IleThrValLysTyrAspAlaLysVal 376

     445 GGGACTAACGGCCATCCTTATGGCGGCGATTATCATATGCCGCGTTTGCA 494
         |||   :::|||       :::   :::|||
     377 Gly...AspGlyLeuLysPheAspSerAsp.................Ly 386

     495 CAAATTTGTCACAGATGCAGAACCTGTTGAGATGACCAGTTATATGGATG 544
         |||:::|||:::|||:::       :::   :::|||          |
     386 sLysIleValAlaAspThrThrAlaLeuThrValThr............G 399

     545 GGTGGAAATACGCTGATTTAAATAAATACCCTGATCGTGTTCGAATC... 591
         ||   |||   |||::::::   |||      |||:::   ::::::
     399 lyGlyLysValAlaGluIleAlaLysGluAspAspLysLysLysLeuVal 415

     592 .....................GGAGCAGGCAGACAATATTGGCGG..... 615
                               :::   |||:::      |||:::
     416 AsnAlaGlyAspLeuValThrAlaLeuGlyAsnLeuSerTrpLysAlaLy 432

     616 .......TCTGATGAAGAC.........GAACCCAATAACCGCGAAAGTT 649
                 :::|||   |||         |||      ::::::::::::::
     432 sAlaGluAlaAspThrAspGlyAlaLeuGluGlyIleSerLysAspGlnG 449

     650 CATATCATATTGCAAGCGCATATTCTTGGCTCGTCGGTGGCAAT...ACC 696
         ::          :::::::::   ::::::   :::|||   |||
     449 luValLysAlaGlyGluThrValThrPheLysAlaGlyLysAsnLeuLys 465

     697 TTTGCACAAAATGGATCAGGT............................. 717
               |||:::|||::::::
     466 ValLysGlnAspGlyAlaAsnPheThrTyrSerLeuGlnAspAlaLeuTh 482

     718 .....................GGTGGCACAGTCAACTTAGGTAGCGAAA 745
                               |||||||||:::|||   |||::::::
     482 rGlyLeuThrSerIleThrLeuGlyGlyThrThrAsnGlyGlyAsnAspA 499

     746 AA.........ATTAAACATAGCCCATATGGTTTTTTACCAACAGGAGGC 786
                   |||:::   :::         :::   |||:::|||:::
     499 laLysThrValIleAsnLysAspGlyLeuThrIleThrProAlaGlyAsn 515

     787 TCATTTGGCGACAGTGGCTCACCAATGTTTATCTATGATGCCCAAAAGCA 836
         :::   |||   :::|||:::      :::
     516 Gly...GlyThrThrGlyThrAsnThrIleSer.................. 525

     837 AAAGTGGTTAATTAATGGGGTATTGCAAACAGGCAACCCCTATATA...G 883
```


526 .....ValThrLysAspGlyIleLysAlaGlyAsnLysAlaIleThrA 540
884 GAAAAAGCAATGGCTTCCAGCTAGTTCGTAAAGATTGGTTCTATGATGAA 933
540 snValAlaSerGlyLeuArgAla.....TyrAspAsp 550
934 ATCTTTGCTGGAGATACCCATTCAGTATTCTACGAACCACATCAAAATGG 983
551 Ala..... 551
984 GAAATACTTTTTTAACGACAATAATAATGGCGCAGGAAAAATCGATGCCA 1033
552 .....AsnPheAspValLeuAsnAsnSerAlaThrAspLeu..... 563
1034 AACATAAACACTATTCTCTACCTTATAGATTAAAAACACGAACCGTTCAA 1083
564 ..AsnArgHisValGluAspAlaTyrLys.....Gly 573
1084 TTGTTTAATGTTTCTTTATCCGAGACAGCAAGAGAACCTGTTTAT..... 1128
574 LeuLeuAsnLeuAsnGluLysAsnAlaAsnLysGlnProLeuValThrAs 590
1129 .CATGCTGCAGGTGGGGTCAACAGTTATCGACCCAGACTGAATAATGGAG 1177
590 pSerThrAlaAlaThrValGlyAspLeuArgLys..... 601
1178 AAAATATTTCCTTTATTGACAAAGGAAAAGGTGAATTGATACTTACCAGC 1227
602 .....LeuGlyTrpValValSerThrLys 609
1228 AACATCAACCAAGGCGCGGGCGGTTTGTATTTTGAGGGTAATTTTACGGT 1277
610 Asn.....GlyThrLysGluGluSerAsn...GlnVa 619
1278 CTCGCCTAAAAACAACGAAACGTGGCAAGGCGCGGGCGTTCATATCAGTG 1327
619 lLysGlnAlaAspGluValLeuPheThrGlyAlaGly..... 631
1328 ATGGCAGTACCGTTACTTGGAAAGTAAACGGCGTGGCAAACGACCGCCTG 1377
632 ..AlaAlaThrValThrSerLys..... 638
1378 TCCAAAATCGGCAAAGGCACGCTGCTGGTTCAAGCCAAAGGGGAAAACCA 1427
639 SerGluAsnGlyLysHisThrIle..... 646
1428 AGGCTCGGTCAGCGTGGGCGACGGTAAAGTCATCTTAGATCAGCAGGCGG 1477
647 ....ThrValSerValAlaGluThrLys..... 654
1478 ACGATCAAGGCAAAAAACAAGCCTTTAGTGAAATCGGCTTGGTCAGCGGC 1527
655 .....AlaAspCysGlyLeuGluLysAsp 662
1528 AGGGGGACGGTGCAACTGAATGCCGATAATCAGTTCAACCCCGACAAACT 1577
663 GlyAspThrIleLysLeuLysValAspAsnGlnAsnThrAspAsnValLe 679
1578 CTATTTCGGCTTTCGCGGCGGACGTTTGGATTTGAACGGGCATTCGCTTT 1627
679 uThrValGlyAsnAsnGlyThrAlaValThrLysGlyGlyPheGluThrV 696
1628 CGTTCCACCGCATTCAAAATACCGATGAAGGGGCGATGATTGTCAAC... 1674
696 alLysThrGlyAlaThrAspAlaAspArgGlyLysValThrValLysAsp 712
1675 .....CACAATCAAGACAAAGAA..... 1692
713 AlaThrAlaAsnAspAlaAspLysLysValAlaThrValLysAspValAl 729
1693 .....TCCACCGTTACCATTACAGGCAATAAAGATATTACTA 1729


729 aThrAlaIleAsnSerAlaAlaThrPheValLysThrGluAsnLeuThrT 746
1730 CAACC...GGCAATAACAACAACTTGGATAGCAAAAAAGAAATTGCCTAC 1776
746 hrSerIleAspGluAspAsnProThrAspAsnGlyLysAspAspAlaLeu 762
1777 AACGGTTGG.....TTTGGCGAGAAAGATGCAACCAA 1808
763 LysAlaGlyAspThrLeuThrPheLysAlaGlyLysAsnLeuLysValLy 779
1809 AACGAACGGGCGG.....CTCAATCTGAATTACC 1837
779 sArgAspGlyLysAsnIleThrPheAspLeuAlaLysAsnLeuGluValL 796
1838 AACCGGAAGAAGCGGATCGCACTTTACTGCTTTCCGGCGGAACAAATTTA 1887
796 ysThrAlaLysValSerAspThrLeuThrIleGlyGlyAsnThrProThr 812
1888 AACGGC.....AATATCACGCAAACAAACGGCAA 1916
813 GlyGlyThrThrAlaThrProLysValAsnIleThrSerThrAlaAspGl 829
1917 ACTGTTTTTCAGCGGCAGACCGACACCGCACGCCTACAATCATTTAGGAA 1966
829 yLeuAsnPheAla...LysGluThrAlaAspAla.....SerGlyS 842
1967 GCGGGTGGTCAAAAATGGAAGGTATCCCACAAGGAGAAATCGTGTGGGAC 2016
842 erLysAsnValTyrLeuLysGlyIleAla..... 851
2017 AACGATTGGATCGACCGCACATTTAAAGCGGAAAACTTCCATATTCAG.. 2064
852 ThrThrLeuThrGluProSerAlaGlyAlaLysSerSerHisValAspLe 868
2065 .GGCGGACAAGCGGTGGTTTCCCGCAATGTTGCCAAAGTGGAA..... 2106
868 uAsnValAspAlaThrLysLysSerAsnAlaAlaSerIleGluAspValL 885
2107 .....GGCGATTGGCATTTAAGCAATCACGCCCAAGCAGTTTTCGGTGTC 2151
885 euArgAlaGlyTrpAsnIleGlnGlyAsnGlyAsnAsnValAspTyrVal 901
2152 GCACCGCATCAAAGCCACACAATCTGTACACGTTCGGACTGGACGGGTCT 2201
902 AlaThrTyrAsp.....ThrValAsnPheThrAspAspSerThrGlyTh 916
2202 GACAAGTTGTACCGAAAAAACCATTACCGACGATAAA.....GTGATTG 2245
916 rThrThrValThrValThrGlnLysAlaAspGlyLysGlyAlaAspValL 933
2246 CTTCATTGAGCAAGACCGACATC..... 2268
933 ysIleGlyAlaLysThrSerValIleLysAspHisAsnGlyLysLeuPhe 949
2269 .....AGAGGCAATGTCAGCCTTGCCGA 2291
950 ThrGlyLysAspLeuLysAspAlaAsnAsnGlyAlaThrValSerGluAs 966
2292 TCACGCTCATTTAAATCTCACAGGACTTGCCACACTCAACGGCAATCTTA 2341
966 pAspGlyLysAspThrGlyThrGlyLeuValThrAlaLysThrValIleA 983
2342 GTGCAGGCGGAGACACGCACTATACGGTTACG...CGCAACGCCACCCAA 2388
983 spAlaValAsnLysSerGlyTrpArgValThrGlyGluGlyAlaThrAla 999
2389 AACGGCAACCTCAGCCTCGTG.....GGCAATGCCCAAGCAACATTTAA 2432
1000 GluThrGlyAlaThrAlaValAsnAlaGlyAsnAlaGluThrValThrSe 1016
2433 TCAAGCCACATTA.....AACGGCAAC.....ACATCGG 2461
1016 rGlyThrSerValAsnPheLysAsnGlyAsnAlaThrThrAlaThrValS 1033



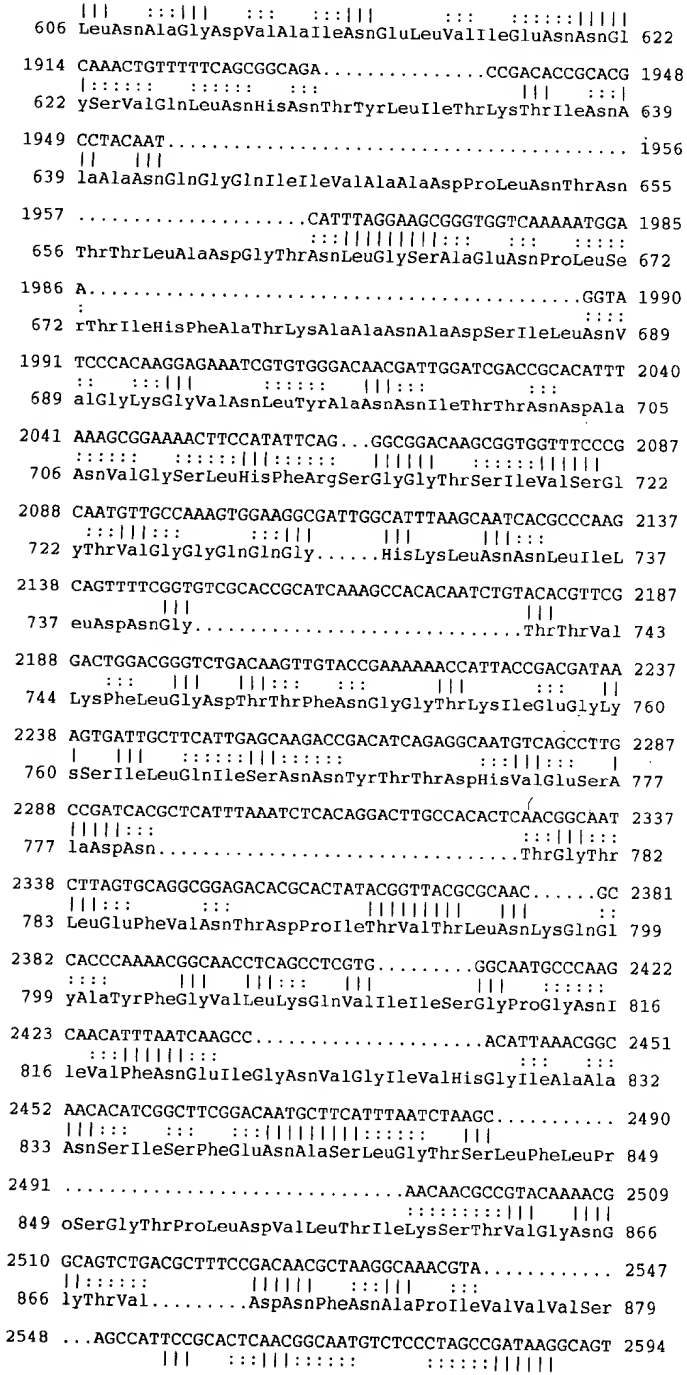
```
                                  :::      :::::::::      ::
    1537 IleAspPheValAlaThrTyrAspThrValAspPheValSerGlyAspLy 1553

    3997 AGATACCGCGCAGGTTTCGGCGGATTCGGCATCGAACCGCACATCGGCGC 4046
         :||||||::::::|||:::::::::|||   :::|||       ::::
    1553 sAspThrThrSerValThrValGluSerLysAspAsnGlyLysArgThrG 1570

    4047 AACGCGCTATTTCGTCCAAAAAGCGGATT 4075
         ::      |||:::::::|||      :::
    1570 luValLysIleGlyAlaLysThrSerVal 1579

seq_name: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-09-413-814-78

seq_documentation_block:
; Sequence 78, Application US/09413814
; Patent No. 6225064
; GENERAL INFORMATION:
;  APPLICANT: Gesellschaft fuer Biotechnologische Forschung mbH
;  APPLICANT: Bristol-Myers Squibb, Co.
;  APPLICANT: Beyer, Stefan
;  APPLICANT: Bloecker, Helmut
;  APPLICANT: Brandt, Petra
;  APPLICANT: Cino, Paul M
;  APPLICANT: Dougherty, Brian A
;  APPLICANT: Goldberg, Steven L
;  APPLICANT: Hofle, Gerhard
;  APPLICANT: Mueller, Joachim
;  APPLICANT: Reichenbach, Hans
;  TITLE OF INVENTION: DNA sequences for enzymatic synthesis of polyketide or
;  TITLE OF INVENTION: heteropolyketide compounds
;  FILE REFERENCE: PCT/US 99/23535
;  CURRENT APPLICATION NUMBER: US/09/413,814
;  CURRENT FILING DATE: 1999-10-07
;  EARLIER APPLICATION NUMBER: DE 198 46 493.2
;  EARLIER FILING DATE: 1998-10-09
;  NUMBER OF SEQ ID NOS: 107
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 78
;   LENGTH: 882
;   TYPE: PRT
;   ORGANISM: Sorangium cellulosum
US-09-413-814-78


alignment_scores:
            Quality:  245.00            Length:    844
              Ratio:  0.714               Gaps:     42
 Percent Similarity:  40.640   Percent Identity: 24.526

alignment_block:
 US-09-303-518D-653 x US-09-413-814-78   ..

 Align seg 1/1  to: US-09-413-814-78  from: 1  to: 882

    2539 GCAAACGTAAGCCATTCC.....................GCACTCAACGG 2567
         |||   ::::::||||||                     ::::::   ||
      75 AlaMetIleGlyHisSerLeuGlyGluLeuValAlaAlaCysValAlaGl 91

    2568 CAATGTCTCCCTAGCCGATAAGGCAGTATTCCATTTTGAAAACAGCCGCT 2617
         |     :::|||   |||          :::   :::::::::::::||||
      91 yValPheThrLeuGluAspAlaMetSerLeuValIleAspArgGlyArgP 108

    2618 TTACCGGAAAAATCAGCGGCGGCAAGGATACGGCATTACACTTA...... 2661
         ||                :::|||:::      |||:::   |||
     108 heMetGlnAlaAlaProAlaGlySerMetLeuAlaValGlyLeuProAla 124

    2662 AAAGACAGCGAATGGACGCTGCCGTCGGGCACGGAATTAGGCAATTTAAA 2711
            |||   |||      |||   :::|||   |||::::::   :::||
      125 AlaAspValGluGlyLeuLeuGluAlaGlyLeuGluIleAlaAlaValAs 141

    2712 CCTTGACAACGCCACCATTACACTCAATTCCG.................. 2743
         |      :::   |||::::::   ::: |||
      141 nSerProLysLeuThrValValAlaGly.ProAlaSerAlaIleArgAsp 157

    2744 .....................................CCTATCGACAC 2754
                                             |||:::::::||
      158 LeuAlaAlaArgLeuGluAlaArgGluValPheAlaArgProLeuGlnTh 174

    2755 GATGCGGCAG.......................GCGCGCAAACCGGCAG 2780
         |   :::                          |||      |||   :
      174 rSerHisAlaPheHisCysAlaLeuIleAspGlyAlaValAlaProPheL 191

    2781 TGCGGCAGATGCGCCGCGCCGCGTTCGCGCCG......TTCCCTATTAT 2824
         ::::::::::|||||||||   ::: :::|||      :::|||:::::
      191 euGluSerValArgArgAlaArgLeu.SerProProGluIleProValVa 207

    2825 C.............................................C 2826
         :
      207 lSerAsnValThrGlyAlaLeuLeuThrAspAlaGluAlaThrAspProA 224

    2827 GTTACGCCGCCAACTTCGGCAGAATCCCGTTTCAACACGCTGACGGTAAA 2876
           |||   |||:::      |||   |||   :::   :::||||||
      224 laTyrTrpAlaArgHisLeuArgGlnProValArgPheSerAspGlyVal 240

    2877 CGGCAAATTGAACGGTCAGGGAAC...ATTCCGCTTTATGTCGGAACT.. 2921
         :::   :::      ||||||:::   :::   |||   ||||||
      241 GluAlaLeuPheAlaSerGlyHisAlaLeuPheLeuGluValGlyProGl 257

    2922 .......CTTCGGCTACCGCAGCGGCAAATTGAAGCTGGCGGAAAGTTCC 2964
                 |||            ||||||      |||||| :::::::::
      257 yArgGlyLeuThrThrLeuValArgGlnThrLeuAlaGly.ArgGlyGly 273

    2965 GAAGGCACTTACACCTTGGCTGTCAACAATACCGGCAACGAACCCGTAAG 3014
         :::      :::|||:::   ::::::::::::::|||||||:::||
      274 AlaAlaIleAlaSerLeuGlySerThrHisAlaAlaSerGluProAlaSe 290

    3015 TCTCGA..................................GCAATTGACG 3030
         ||||                                   |||:::|||
      290 rLeuAlaGluAlaLeuGlyGlnLeuTrpGluAlaGlyHisAlaValAspT 307

    3031 GTAGTGGAAGGAAAAGACAACACACCGCTGTCCGAAAATCTTAATTTCAC 3080
           ::::::|||   |||   :::      :::|||:::|||
      307 rpThrAlaArgProArgGlyArgProProAlaArgSerSer......... 320

    3081 CCTGCAAAACGAACACGTCGA.................TGCCGGCGCAT 3112
         ::::::||||||||||                      ||||||
      321 AlaCysArgArgThrArgSerSerGlyArgGlyThrGlySerArgArgAl 337

    3113 GGCGTTATCAGCTTATCCGCAAAGACGGCGAGTTCCGCCTGCATAATCCG 3162
         |||          |||:::||||||||||   |||          |||
      337 aAlaAlaProProProProSerArgArgArg...ProThrArgProSerA 353

    3163 GTCAAAGAACAAGAGCTTTCCGACAAACTCGGCAAGGCGGGAGAAACAGA 3212
            |||             |||:::   |||:::   :::|||:::
      353 rgProArgArgThrProSerArgArgArgArgSerProAlaArgThrPro 369

    3213 GGCCGCCTTGACGGCAAAACAGGCACAACTTGCCG.............. 3247
         |||  |||          :::|||:::      |||
      370 Gly..ProArgSerProArgArgArgProSerProSerAlaArgSerPro 385

    3248 ........................CCAAACAACAGGCGGAAAAAGAC 3270
                                 |||:::::::||||||   :::
      386 AlaIleTrpGluArgProProArgArgProArgAspArgArgProArgAr 402

    3271 AACGCGCAAAGCCTTGACGCGCTGATTGCGG......CCGGGCGCAATGC 3314
            :::         |||      |||      ||||||   :::|
      402 gLeuLeuArgProAlaArgArgLeuAlaArgArgAspProGlyAspValP 419

    3315 CACCGAAAAGGCAGAAAGTGTTGCCGAACCGGCCCGGCAGGCAGGCGGGG 3364
         ||      |||              :::|||   :::      ||||||
      419 roAspProArg.........AlaAlaArgArgArgAlaProAlaAlaGly 432

    3365 AAAATGCCGGCATTATGCAGGCGGAGGAAGAGAAAAAACGGGTGCAGGCG 3414
            :::|||:::         |||      |||      |||   |||||
      433 AlaLeuProGlyAlaAspGlyArgGlyAlaArgGlyProGlyArgArgAr 449
```

```
3415 GATAAAGACACCGCCTTGGCGAAACAGCGCGAAGCGGAAACCCGGCCGGC 3464
     |          |||   |||:::::::|||           :::|||:
 449 g.........AlaProArgArgGlyAlaAla...........AlaArgV 459

3465 TACCACCGCCTTCCCCCGCGCCCGCCGCGCCCGCCGGGATTTGCCGCAAC 3514
     ::|||||||             |||||||||   |||:::   |||:::::
 459 alProProAspArg.AlaAspProProArgArgProAlaProAlaValVa 475

3515 C...GCAGCCCCAACCGCAACCCCAACCGCAGCGCGACCTGATCAGCCGT 3561
     :   ||||||||||:::::::::      :::|||||||      ||||||
 475 lArgAlaAlaProAlaValValArgArgSerAlaArgAlaGlyGlnPro. 491

3562 TATGCCAATAGCGGTTTGAGTGAATTTTCCGCCACGCTCAACAGCGTTTT 3611
                                        :::   |||:::
 492 ..................................GlyValGlnHisPro 496

3612 CGCCGTACAGGACGAATTGGACCG...CGTGTTTGCCGAAGACCGCCGCA 3658
        |||   :::|||   ||||||   |||      |||   ||||||||
 497 ValArgGlyAlaArgAspGlyProAlaArgArgArgArgAlaProProGl 513

3659 ACGCCGTTTGGACAAGCGGCATCCG........................... 3683
     |        |||:::      |||
 513 nProValArgAspArgAlaAlaProArgGlyAlaAlaAspAspValGlnA 530

3684 .....GGACACCAAACACTACCGTTCGCAAGATTT.............. 3713
          |||:::::::::::::|||      |||
 530 rgGlnGlyArgArgAlaValProGlyArgArgAlaArgGlyProGlyAla 546

3714 ..............CCGCGCCTACCG..................... 3725
                    ||||||  |||
 547 ValProAspValArgProArgAlaProArgGlyArgGlyProArgArgGl 563

3726 ........................................CCAACAA 3732
                                             |||
 563 yGlyLeuGlyAlaArgProArgGlyGluProArgAlaValArgProGluP 580

3733 ACCGACCTGCGCCAAATCGG......TATGCAGAAAAACCTCGGCAG... 3773
       |||   ||||||   |||      :::   |||   |||||||::
 580 roArgAlaAlaProProArgAlaGlyHisProGluAlaProArgArgAla 596

3774 ......CGGGCGCGTCGGCATCCT.............GTTTTCGCACA 3802
           ||||||||||||   |||              :::   :::::
 597 ArgHisArgAlaArgArgAlaProArgArgValArgArgLeuValGlyAr 613

3803 ACCGGACCGGAAACACCTTCGACGACGGCATCGGCAACTC......GGCA 3846
     :      |||::::::   ||||||      |||:::|||      |||
 613 gArgLeuArgArgArgAlaArgArgAlaLeuArgArgLeuArgAlaGlyP 630

3847 CGGCTTGCCCACGGTGC............CGTTTTCGGGCAATACGGCAT 3884
       |||   |||   :::                |||   :::|||:::
 630 roAlaPheProAlaAlaGlyAlaProGlyAlaValArgArgLeuArgArg 646

3885 CGGCAGGTTCGACATCGGCATCAGCGCGGGCGCGGGTTTTAGTAGCGGCA 3934
        :::                :::||||||   |||          |||
 647 SerProAlaGlyValAlaValArgArgGlyProGlyGly......ArgAl 661

3935 GCCTTTC..................................... 3941
      |||
 661 aProValLeuAspAspGluAlaLeuGlyArgAlaAlaProGlyProArgA 678

3941 ..................................................... 3941

 678 laGlyArgProArgAlaGlyGlyAlaLysAspValAlaArgGlyAlaAla 694

3942 ......AGACGGCATCAGAGGCAAAATCCGCCGCCGCGTGCTGCATTACG 3985
           ||||||:::   |||   :::|||||||   :::::::::|||||
 695 GluProArgArgArgGlyArgAlaHisProProAspGlnGlyValLeuAr 711

3986 GCATTCAGGCAAGATACCGCGCAGG......TTTCGGCGGATTCGGCATC 4029
     |   :::|||   :::|||   :::         ||||||::::::   |
 711 gAlaGlyGlyAspLeuProLeuHisGlyProAlaArgArgValGlnAlaA 728

4030 GAACCGCA..............................CATCGGCGC 4046

        ||   |||                                  |||||||||
 728 rgProAlaProAlaHisArgAlaArgGlyProArgArgGlyHisArgArg 744

4047 ................................AACGCGCTATTTCGTCC 4063
                                     :::   |||:::|||
 745 ArgGluProGluProArgArgAspArgAlaAspAspArgLeuLeuArgGl 761

4064 AAAAAGCGGATTACCG................................ 4079
          :::      |||
 761 nProAlaArgAlaProAspArgLeuArgGlyArgProAspValArgArgA 778

4080 ...................ATACGAAAACGTCAATATCGCCACCCCGGG 4109
                                 |||   |||      |||   ||||||
 778 laGlyProProArgAlaArgArgArgAlaArgGlyValArgAlaProGly 794

4110 CCTTGCATTCAACCGCTACCGCGCGGGCATTAAGGCAGATTATTCATTCA 4159
     |||::::::::::|||      ||||||
 795 ProAlaValArgProAlaArgArgGlyPro..............AlaAl 806

4160 AACCGGCGCAACACATTTCCATCACGCCTTATTTGAGCCTGTCCTATACC 4209
          |||:::   :::   :::      |||   :::   |||   :::
 806 aGluGlyGlyArgArgAlaArgAlaProLeuArgArgGluValArgHisA 823

4210 GATGCCGCTTCCGGCAAAGTCCGAACGCGCGTCAATACCGCCGTATTGGC 4259
          |||   |||         :::||||||      |||      |||
 823 laGlnArgAlaArgProProAspGluAlaArgGlyAlaArgAlaArgGly 839

4260 GCA.....................GGATTTCGGCAAAACCCGCAGTGCGG 4288
                             |||   |||   :::|||:::
 840 AlaGlyGlyArgGlyAspHisAspGlyValArgLeuArgProHisGlyAr 856

4289 AATGGGGCGTAAA............CGCCGAAATCAAAGGTTTCACGCTG 4326
          ||||||:::            ||||||:::   :::   :::||||
 856 gArgGlyArgArgLeuValProLeuArgArgArgAlaGlnGlnArgAlaV 873

4327 TCCCTCCACGCTGCCGCCGCCAA 4349
     |||||          ||||||
 873 alProGlyGlyAspGlyArgGln 880
```

seq_name: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:US-08-169-927-2

```
seq_documentation_block:
; Sequence 2, Application US/08169927
; Patent No. 5783441
;  GENERAL INFORMATION:
;    APPLICANT:  Carl, Mitchell
;    APPLICANT:  Dobson, Michael E.
;    APPLICANT:  Ching, Wei Mei
;    APPLICANT:  Dasch, Gregory A
;    TITLE OF INVENTION:  Gene and Protein Applicable to the
;    TITLE OF INVENTION:  Preparation of Vaccines for Rickettsia prowazekii and
;    TITLE OF INVENTION:  Rickettsia typhi and the Detection of Both
;    NUMBER OF SEQUENCES:  2
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Counsel, Naval Medical R & D Command
;      STREET:  Bldg. 1, T-12, 8901 Wisconsin Ave.
;      CITY:  Bethesda
;      STATE:  MD
;      COUNTRY:  USA
;      ZIP:  20889-5606
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/169,927
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/742,128
;      FILING DATE:  08/09/91
;    ATTORNEY/AGENT INFORMATION:
```

```
;      NAME:  Spevack, A. David
;      REGISTRATION NUMBER:  24,743
;      REFERENCE/DOCKET NUMBER:  75,976
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (301) 295-6759
;      TELEFAX:  (301) 295-1022
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1612 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-169-927-2


alignment_scores:
            Quality:  243.50              Length:   1471
              Ratio:  0.362                Gaps:     72
 Percent Similarity:  45.683   Percent Identity: 19.511

alignment_block:
 US-09-303-518D-653 x US-08-169-927-2   ..

 Align seg 1/1  to: US-08-169-927-2  from: 1  to: 1612

     634 AATAACCGCGAAAGTTCATATCATATTGCAAGCGCATATTCTTGGCTCGT 683
         |||:::|||:::||||||||||||||:::::|||
     160 AsnSerArgSerSerSerTyrHisLeuValSer..................  170

     684 CGGTGGCAATACCTTTGCACAAAATGGATCAGGTGGTGGCACAGTCAACT 733
               |||:::         ||||||   |||:::|||   :::||||:
     171 .......AsnSerLysIleValAsnGly...GlyAsnGlyIleLeuAsnI 184

     734 TA......GGTAGCGAAAAAATTAAACATAGCCCATATGGTTTTTTACCA 777
         ::      |||      :::::::::   :::   ::::::   :::
     184 leThrAsnGlyPheIleGlnValSerAspAsnThrPheAlaGlyIleLys 200

     778 ACAGGAGGCTCATTTGGCGACAGTGGCTCACCAATGTTTATCTATGATGC 827
         |||      ::::::   |||      :::   ||||||
     201 Thr...IleAsnIleAspAspCysGlnGlyLeuMetPheAsnSerThrPr 216

     828 CCAAAAGCAAAAGTGGTTA............................A 847
          :::      :::   |||                          :
     216 oAspAlaAlaAsnThrLeuAsnLeuGlnAlaGlyGlyAsnThrIleAsnP 233

     848 TTAATGGGGTATTGCAAACAGGCAACCCCTATATAGGAAAAAGCAATGGC 897
         ::|||||:::      ||||||:::      :::   ::::::||||||
     233 heAsnGlyIleAspGlyThrGlyLysLeuValLeuValSerLysAsnGly 249

     898 ............TTCCAGCTAGTTCGTAAAGATTGGTTCTATGATGAAAT 935
                     |||:::::::::       :::|||:::  :::|||:
     250 AlaAlaThrGluPheAsnValThrGly.ThrLeuGlyGlyAsnLeuLysG 266

     936 CTTTGCTGGAGATA...CCCATTCAGTATTCTACGAACCACATCAAAATG 982
         ::::::::|||:::              :::   :::   |||::::::
     266 lyIleIleGluLeuAsnThrAlaAlaValAlaGlyLysLeuIleSerLeu 282

     983 GGA.........AATACTTTTTTAACGACAATAATAATGGCGCAGGAAAA 1023
         |||         |||:::   :::   |||  :::||||||||||||:::
     283 GlyGlyAlaAlaAsnAlaValIleGlyThr..AspAsnGlyAlaGlyArg 298

    1024 ATCGATGCCAAACATAAACACTATTCTCTACCTTATAGATTAAAAACACG 1073
            :::
     299 AlaAlaGly.......................................... 301

    1074 AACCGTTCAATTGTTTAATGTTTCTTTATCCGAGACAGCAAGAGAACCTG 1123
                  |||  ||||||::::::::                   :
     302 ............PheIleValSerValAspAsnGlyAsnAlaAlaThrI 314

    1124 TTTATCATGCTGCAGGTGGGGTCAACAGTTATCGACCCAGACTGAATAAT 1173
         ::         :::   :::   |||                 |||
     314 leSerGlyGlnValTyrAlaLysAsnMetValIleGlnSerAlaAsnAla 330

    1174 GGAGAAAATATTTCCTTT.........ATTGACAAAGGAAAAGGT..... 1209
         |||   ::::::::::|||         :::|||   |||   |||
     331 GlyGlyGlnValThrPheGluHisIleValAspValGlyLeuGlyGlyTh 347

    1210 ........................GAATTGATACTTACCAGCAACATCA 1234
                                 :::::::|||:::|||:::|||   |
     347 rThrAsnPheLysThrAlaAspSerLysValIleIleThrGluAsnSerA 364

    1235 ACCAAGGCGCG.........GGCGGTTTG..................... 1254
         ||   |||:::          |||:::|||
     364 snPheGlySerThrAsnPheGlyAsnLeuAspThrGlnIleValValPro 380

    1255 .........TATTTTGAGGGTAATTTTACGGTCTCGCCTAAAAACAACGA 1295
                   ::::::||||||||||      :::   |||||||||
     381 AspThrLysIleLeuLysGlyAsnPheIleGlyAspValLysAsnAsnGl 397

    1296 A.................ACGTGGCAAGGCGCGGGCGTTCATATCAGT. 1326
                           |||::::::::::   |||:::   :::|||
     397 yAsnThrAlaGlyValIleThrPheAsnAlaAsnGlyAlaLeuValSerA 414

    1327 ........GATGGCAGTACCGTTACTTGGAAAGTAAACGGC......... 1359
                 |||   :::   ::::::   ::::::|||:::
     414 laSerThrAspProAsnIleAlaValThrAsnIleAsnAlaIleGluAla 430

    1360 ................................GTGGCAAACGACCG 1373
                                        :::|||:::   ||
     431 GluGlyAlaGlyValValGluLeuSerGlyIleHisIleAlaGluLeuAr 447

    1374 CCTGTCC.................AAAATCGGCAAAGGCACGCTGCTGG 1405
         ||||:::                |||::::::   ||||||::::::
     447 gLeuGlyAsnGlyGlySerIlePheLysLeuAlaAspGlyThrValIleA 464

    1406 TTCAAGCCAAAGGGGAAAAC.........................CAA 1428
                        ::::::|||
     464 snGlyProValAsnGlnAsnAlaLeuMetAsnAsnAsnAlaLeuAlaAla 480

    1429 GGCTCGGTCAGCGTGGGCGACGGTAAAGTCATCTTAGATCAGCAGGCGGA 1478
         ||||||::::::::::   ||||||:::::|||:::      :::
     481 GlySerIleGlnLeu...AspGlySerAlaIleIleThrGlyAspIleGl 496

    1479 CGATCAAGGCAAAAAACAAGCCTTTAGTGAAATCGGCTTGGTCAGCGGCA 1528
          :::   |||   :::   |||::::::::::|||   |||::::::
     496 yAsnGlyGlyValAsnAlaAlaLeuGlnHisIleThrLeuAlaAsnAspA 513

    1529 GGGGGACGGTG...CAACTGAATGCCGATAATCAGTTCAACCCCGAC... 1572
           :::   :::      |||::::::   |||   ::::::   :::
     513 laSerLysIleLeuAlaLeuAspGlyAlaAsnIleIleGlyAlaAsnVal 529

    1573 ......AAACTCTATTTCGGCTTTCGCGGCGGACGTTTGGATTTGAACGG 1616
               :::::::|||      :::||||||   :::
     530 GlyGlyAlaIleHisPheGlnAlaAsnGlyGlyThrIle........... 542

    1617 GCATTCGCTTTCGTTCCACCGCATTCAAAATACCGATGAAGGGGCGATGA 1666
                         ::::::   ||||||::::::::::   ::::
     543 ...................LysLeuThrAsnThrGlnAsnAsnIleValV 553

    1667 TTGTCAACCACAATCAAGACAAAGAATCCACCGTTACCATTACAGGCAAT 1716
         ::         |||   |||   :::   |||:::      ||||||
     553 al.........AsnPheAspLeuAspIleThrThrAspLysThrGlyVal 566

    1717 AAAGATATTACTACAACCGGCAATAACAACAACTTGGATAGCAAAAAAGA 1766
            |||   ::::::      ||||||::::::|||
     567 ValAspAlaSerSerLeuThrAsnAsnGlnThrLeu.............. 578

    1767 AATTGCCTACAACGGTTGGTTTGGCGAGAAAGATGCA...ACCAAAACGA 1813
            :::   ||||||   :::|||         |||   |||||||||
     579 ....ThrIleAsnGlySerIleGlyThrValValAlaAsnThrLysThrL 594

    1814 ACGGGCGGCTCAATCTGAATTACCAACCGGAAGAAGCGGATCGCACTTTA 1863
         ::::::||||||::::::              :::::::::|||:::
     594 euAlaGlnLeuAsnIleGly..............SerSerLysThrIle 605

    1864 CTGCTTTCCGGCGGAACAAATTTAAACGGCAATATCACGCAAACAAACGG 1913
```


606 LeuAsnAlaGlyAspValAlaIleAsnGluLeuValIleGluAsnAsnGl 622
1914 CAAACTGTTTTTCAGCGGCAGA.....CCGACACCGCACG 1948
622 ySerValGlnLeuAsnHisAsnThrTyrLeuIleThrLysThrIleAsnA 639
1949 CCTACAAT..... 1956
639 laAlaAsnGlnGlyGlnIleIleValAlaAlaAspProLeuAsnThrAsn 655
1957 .....CATTTAGGAAGCGGGTGGTCAAAAATGGA 1985
656 ThrThrLeuAlaAspGlyThrAsnLeuGlySerAlaGluAsnProLeuSe 672
1986 A.....GGTA 1990
672 rThrIleHisPheAlaThrLysAlaAlaAsnAlaAspSerIleLeuAsnV 689
1991 TCCCACAAGGAGAAATCGTGTGGGACAACGATTGGATCGACCGCACATTT 2040
689 alGlyLysGlyValAsnLeuTyrAlaAsnAsnIleThrThrAsnAspAla 705
2041 AAAGCGGAAAACTTCCATATTCAG...GGCGGACAAGCGGTGGTTTCCCG 2087
706 AsnValGlySerLeuHisPheArgSerGlyGlyThrSerIleValSerGl 722
2088 CAATGTTGCCAAAGTGGAAGGCGATTGGCATTTAAGCAATCACGCCCAAG 2137
722 yThrValGlyGlyGlnGlnGly......HisLysLeuAsnAsnLeuIleL 737
2138 CAGTTTTCGGTGTCGCACCGCATCAAAGCCACACAATCTGTACACGTTCG 2187
737 euAspAsnGly.....ThrThrVal 743
2188 GACTGGACGGGTCTGACAAGTTGTACCGAAAAAACCATTACCGACGATAA 2237
744 LysPheLeuGlyAspThrThrPheAsnGlyGlyThrLysIleGluGlyLy 760
2238 AGTGATTGCTTCATTGAGCAAGACCGACATCAGAGGCAATGTCAGCCTTG 2287
760 sSerIleLeuGlnIleSerAsnAsnTyrThrThrAspHisValGluSerA 777
2288 CCGATCACGCTCATTTAAATCTCACAGGACTTGCCACACTCAACGGCAAT 2337
777 laAspAsn.....ThrGlyThr 782
2338 CTTAGTGCAGGCGGAGACACGCACTATACGGTTACGCGCAAC......GC 2381
783 LeuGluPheValAsnThrAspProIleThrValThrLeuAsnLysGlnGl 799
2382 CACCCAAAACGGCAACCTCAGCCTCGTG.....GGCAATGCCCAAG 2422
799 yAlaTyrPheGlyValLeuLysGlnValIleIleSerGlyProGlyAsnI 816
2423 CAACATTTAATCAAGCC.....ACATTAAACGGC 2451
816 leValPheAsnGluIleGlyAsnValGlyIleValHisGlyIleAlaAla 832
2452 AACACATCGGCTTCGGACAATGCTTCATTTAATCTAAGC..... 2490
833 AsnSerIleSerPheGluAsnAlaSerLeuGlyThrSerLeuPheLeuPr 849
2491 .....AACAACGCCGTACAAAACG 2509
849 oSerGlyThrProLeuAspValLeuThrIleLysSerThrValGlyAsnG 866
2510 GCAGTCTGACGCTTTCCGACAACGCTAAGGCAAACGTA..... 2547
866 lyThrVal.....AspAsnPheAsnAlaProIleValValValSer 879
2548 ...AGCCATTCCGCACTCAACGGCAATGTCTCCCTAGCCGATAAGGCAGT 2594
880 GlyIleAspSerMetIleAsnAsnGlyGlnIleIleGlyAspLys..... 894
2595 ATTCCATTTTGAAAACAGCCGCTTTACCGGAAAAATCAGCGGCGGCAAGG 2644
895 .....LysAsnIleIleAlaLeuSerLeuGlySerA 905
2645 ATACGGCATTACACTTAAAAGACAGCGAATGGACGCTGCCGTCGGGCACG 2694
905 spAsnSerIleThrValAsnAlaAsn......ThrLeuTyrSerGlyIle 919
2695 GAATTAGGCAATTTAAACCTTGACAACGCCACCATTACACTCAATTCCGC 2744
920 ArgThrThrLysAsnAsn......GlnGlyThrValThrLeuSerGlyGl 934
2745 CTATCGACACGAT.....GCGGCAGGCGCGCAAACCGGCA 2779
934 yMetProAsnAsnProGlyThrIleTyrGlyLeuGlyLeuGluAsnGlyS 951
2780 GTGCGGCAGATGCGCCGCGCCGCCGTTCGCGCCGTTCCCTATTATCCGTT 2829
951 erPro.....LysLeuLysGlnVal 957
2830 ACGCCGCCAACTTCGGCAGAATCCCGT.....TTCAACACGCT 2867
958 ThrPheThrThrAspTyrAsnAsnLeuGlySerIleIleAlaAsnAsnVa 974
2868 GACGGTAAAC..... 2877
974 lThrIleAsnAspAspValThrLeuThrThrGlyGlyIleAlaGlyThrA 991
2878 .....GGCAAATTGAACGGTCAGGGAACATTC 2904
991 spPheAspAlaLysIleThrLeuGlySerValAsnGlyAsnAlaAsnVal 1007
2905 CGCTTTATGTCGGAACTCTTCGGCTACCGCAGCGGCAAATTGAAGCTGGC 2954
1008 ArgPheValAspSerThrPheSerAspProArgSerMetIleValAlaTh 1024
2955 GGAAAGTTCCGAAGGCACTTACACC.....TTGGCTGTCAACA 2992
1024 rGlnAlaAsnLysGlyThrValThrTyrLeuGlyAsnAlaLeuValSerA 1041
2993 ATACCGGC......AACGAACCCGTAAGTCTCGAGCAATTGACGGTAGTG 3036
1041 snIleGlySerLeuAspThrProValAlaSerValArgPheThr..... 1055
3037 GAAGGAAAAGACAACACACCGCTGTCCGAAAATCTTAATTTCACCCTGCA 3086
1056 ...GlyAsnAspSer.....GlyAlaGlyLeuGlnGlyAsnIleTy 1068
3087 AAACGAACACGTCGATGCCGGCGCATGGCGTTATCAGCTTATCCGCAAAG 3136
1068 rSerGlnAsnIleAspPheGlyThrTyrAsnLeuThrIle..... 1081
3137 ACGGCGAGTTCCGCCTGCATAATCCGGTCAAAGAACAAGAGCTTTCCGAC 3186
1082 .....LeuAsnSer 1084
3187 AAACTCGGCAAGGCGGGAGAAACAGAGGCCGCCTTGACGGCAAAACAGGC 3236
1085 AsnValIleLeuGlyGlyGlyThr.....ThrAlaIleAsnGl 1097
3237 ACAACTTGCCGCCAAACAACAGGCGGAAAAAGACAACGCGCAAAGCCTTG 3286
1097 yGluIle.....AspLeuLeuThrAsnA 1105
3287 ACGCGCTGATTGCGGCCGGGCGCAATGCCACCGAAAAGGCAGAAAGTGTT 3336
1105 snLeuIlePheAlaAsnGlyThrSerThrTrpGlyAspAsnThrSerIle 1121
3337 GCCGAACCGGCCCGGCAGGCAGGCGGGGAAAATGCCGGCATTATGCAGGC 3386
1122 SerThrThrLeuAsnValSerSerGlyAsnIleGlyGlnValValIleAl 1138

```
3387 GGAGGAAGAGAAAAAACGGGTGCAGGCGGATAAAGACACCGCCTTGGCGA 3436
1138 aGluAspAlaGlnValAsnAlaThrThrThrGlyThrThrThrIleLysI 1155

3437 AACAGCGCGAAGCGGAAACCCGG...CCGGCTACCACCGCCTTCCCCCGC 3483
1155 leGlnAspAsnAlaAsnAlaAsnPheSerGlyThrGlnAlaTyrThrLeu 1171

3484 GCCCGCCGCGCCCGCCGGGATTTGCCGCAACCGCAGCCCCAACCGCAACC 3533
1172 IleGlnGlyGlyAlaArg..............................  1177

3534 CCAACCGCAGCGCGACCTGATCAGCCGTTATGCCAATAGCGGTTTGAGTG 3583
1178 .......................PheAsnGlyThrLeuGlyAlaProA 1186

3584 AATTTTCCGCCACGCTCAACAGCGTTTTCGCCGTACAGGACGAATTGGAC 3633
1186 snPheAlaValThrGlySerAsnIlePhe...ValLysTyrGluLeu... 1200

3634 CGCGTGTTTGCCGAAGACCGCCGCAACGCCGTTTGGACAAGCGGCATCCG 3683
1201 .............................................IleAr 1202

3684 GGACACCAAACACTACCGTTCGCAAGATTTCCGCGCCTACCGCCAACAAA 3733
1202 gAspSer............AsnGlnAspTyr.................... 1208

3734 CCGACCTGCGCCAAATCGGTATGCAGAAAAACCTCGGCAGCGGGCGCGTC 3783
1208 ..................................................  1208

3784 GGCATCCTGTTTTCGCACAACCGGACCGGAAACACCTTCGACGACGGCAT 3833
1209 ...ValLeuThrArgThrAsnAspValLeuAsnValValThrThrAlaVa 1224

3834 CGGCAACTCGGCACGGCTTGCCCACGGTGCCGTTTTCGGGCAATACGGCA 3883
1224 lGlyAsnSerAlaIleAlaAsnAlaProGlyVal.SerGlnAsnIleSer 1240

3884 TCGGCA...GGTTCGACATCGGCATCAGCGCGGGCGCGGGTTTTAGTAGC 3930
1241 ArgCysLeuGluSerThrAsnThrAlaAlaTyrAsnAsnMetLeuLeuAl 1257

3931 G..................GCAGCCTTTCAGACGGCATCAGAGGCAAAAT 3962
1257 aLysAspProSerAspValAlaThrPheValGlyAlaIleAlaThrAspT 1274

3963 CCGCCGCCGCGTGCTGCATTACGGCATTCAGGCAAGATACCGCGCAGGTT 4012
1274 hrSerAlaAla......ValThrThrValAsnLeuAsnAspThrGlnLys 1288

4013 TCGGCGGATTCGGCATCGAACCGCACATCGGCGCAACGCGCTATTTCGTC 4062
1289 ThrGlnAspLeuLeuSerAsnArg..LeuGlyThrLeuArgTyrLeuSer 1304

4063 CAAAAAGCGGATTACCGATACGAAAACGTCAATATCGCCACCCCGGGCCT 4112
1305 AsnAla..............GluThrSerAspValAla......GlySe 1314

4113 TGCATTCAACCGCTACCGCGCGGGCATTAAGGCAGATTATTCATTCAAAC 4162
1314 rAlaThrGlyAlaValSerSerGlyAspGluAlaGluValSerTyr.... 1329

4163 CGGCGCAACACATTTCCATCACGCCTTATTTGAGCCTGTCCTATACCGAT 4212
1330 ........GlyValTrpAlaLysProPheTyrAsnIleAlaGluGlnAsp 1343

4213 GCCGCTTCCGGC............AAAGTCCGAACGCGC........... 4239
1344 LysLysGlyGlyIleAlaGlyTyrLysAlaLysThrThrGlyValValVa 1360

4239 ..................................................  4239

1360 lGlyLeuAspThrLeuAlaSerAspAsnLeuMetIleGlyAlaAlaIleG 1377

4240 .....GTCAATACCGCCGTATTGGCGCAGGATTTCGGCAAAACCCGCAGT 4284
1377 lyIleThrLysThrAspIleLysHisGlnAspTyrLysLys......... 1390

4285 GCGGAATGGGGCGTAAACGCCGAAATCAAAGGTTTCACGCTGTCCCTCCA 4334
1391 .........GlyAspLysThrAspIleAsnGlyLeuSerPheSerLeuTy 1404

4335 CGCTGCC 4341
1404 rGlySer 1406
```

seq_name: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-09-268-347-49

seq_documentation_block:
; Sequence 49, Application US/09268347
; Patent No. 6335182
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena M.
; TITLE OF INVENTION: RECOMBINANT HAEMOPHILUS INFLUENZAE ADHESIN PROTEINS
; FILE REFERENCE: 1038-860
; CURRENT APPLICATION NUMBER: US/09/268,347
; CURRENT FILING DATE: 1999-03-16
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 49
; LENGTH: 2314
; TYPE: PRT
; ORGANISM: Haemophilus influenzae
US-09-268-347-49

alignment_scores:
Quality: 237.00 Length: 1232
Ratio: 0.422 Gaps: 59
Percent Similarity: 45.536 Percent Identity: 19.805

alignment_block:
US-09-303-518D-653 x US-09-268-347-49 ..

Align seg 1/1 to: US-09-268-347-49 from: 1 to: 2314

```
340 GCACATAACGGCGGCTATAACAATGTTGATTTTGGTGCGGAGGGAAGCAA 389
 25 SerHisSerGlyGlySerSerSerSerThrAlaGlyGlnValGlySerSe 41

390 TCCCGATCAGCACCGCTTTTCTTACCAAATTGTGAAAAGAAATAATTATA 439
 41 r.ProValIleArg..................................... 45

440 AAGCAGGGACTAACGGCCATCCTTATGGCGGCGATTATCATATGCCGCGT 489
 46 .........LeuThrArgValAlaThrLeuAlaIleLeuValIleGlyAl 59

490 TTGCACAAATTTGTCA...CAGATGCAGAACCTGTTGAGATGACCAGTTA 536
 59 aThrLeuAsnGlySerAlaTyrAlaGlnAsnAsnSerLysIleAlaPheG 76

537 TATGGATGGGTGGAAAT.......................ACGCTGATT 562
 76 lyThrThrGlyAsnAsnAspAsnAlaSerAlaSerAsnGluAlaSerIle 92

563 TAAATAAATACCCTGATCGTGTTCGAATCGGAGCAGGCAGACAATATTGG 612
 93 AlaIleGlySerLeuAlaLysAlaHisAlaAsnGlnAlaIleAlaIleGl 109

613 CGGTCTGATGAAGACGAACCCAAT.............AACCGCGAAAGTTC 650
109 yGly..SerLysProAspProArgAsnGlnAlaAlaAsnGlnLysAlaGl 125
```

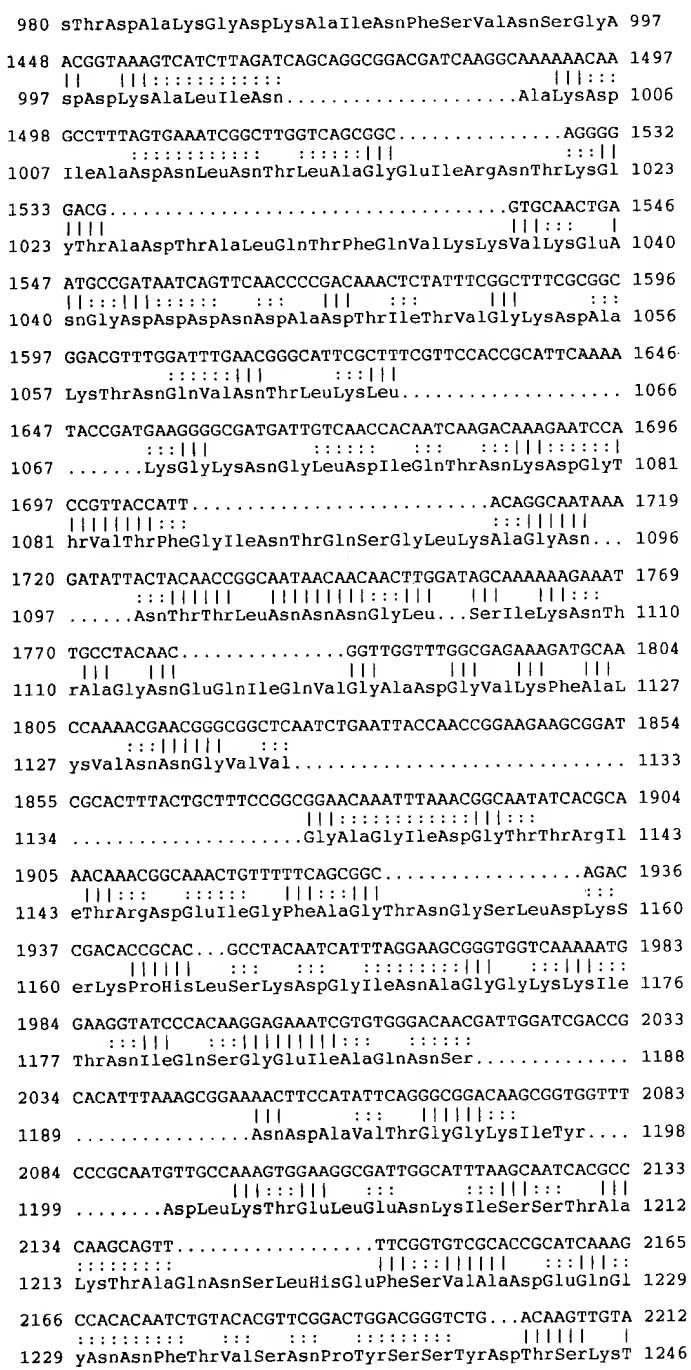
```
 651 ATATCATATTGCAAGCGCATATTCTTGGCTCGTCGGTGGCAATACCTTTG 700
 125 ySerHisAlaLysGlyLysGluSerIleAlaIleGlyGlyAspValLeuA 142

 701 CACAAAATGGATCAGGTGGTGGCACAGTCAACTTAGGTAGCGAAAAAATT 750
 142 laGlu.........GlyAspAlaSerIleAlaIleGlySerAspAspLeu 155

 751 .........AAACATAGCCCATATGGTTTTTTACCAACAGGAGGCTCATT 791
 156 TyrLeuAspArgAsnSerThrAsnSerLysTyrProAsnGly......... 169

 792 TGGCGACAGTGGCTCACCAATGTTTATCTATGATGCCCAAAAGCAAAAGT 841
 169 ..................................................... 169

 842 GGTTAATTAATGGGGTATTGCAA......ACAGGCAACCCCTATATAGGA 885
 170 ..LeuLeuSerThrLeuIleGlnAsnHisThrValLeuArgGlnIleArg 185

 886 AAAAGCAATGGCTTCCAGCTAGTTCGTAAAGATTGGTTCTATGATGAAAT 935
 186 AspSerAsnGlySerGlnLysTyrArgArgThr................... 196

 936 CTTTGCTGGAGATACCCAT.........TCAGTATTCTACGAACCACATC 976
 197 ....AlaAlaGluGlyHisAlaSerThrAlaValGlyAlaMetAlaTyrA 212

 977 AAAATGGGAAATACTTTTTTAACGACAATAATAATGGCGCAGGAAAAATC 1026
 212 laLysGlyHisPhe..............AlaAsnAlaPheGlyThrArg 223

1027 GATGCCAAACATAAACACTATTCTCTACCTTATAGATTAAAAACACGAAC 1076
 224 SerThrAlaGluGlyAsnTyrSerLeuAlaValGlyLeuThrAlaLysAl 240

1077 CGTTCAATTGTTTAATGTTTCTTTATCCGAGACAGCAAGAGAACCTGTTT 1126
 240 aGluLysGlyTyrThrIleAlaIleGlySerAsnAlaGlnAlaIleAsnT 257

1127 ATCATGCTGCA.....................................GGT 1140
 257 yrGlyAlaLeuAlaLeuGlyAlaAspThrArgValAspLeuAspTyrGly 273

1141 GGGGTCAACAGTTATCGACCCAGACTG.............AATAATGGAGA 1178
 274 IleAlaLeuGlyTyrGlySerGlnIleLeuAsnAsnAsnAsnAsnAsAs 290

1179 AAATATTTCCTTTATTGACAAAGGAAAAGGTGAATTGATACTTACCAGCA 1228
 290 nAsnLysAlaTyrValProGluGlyAsnGlySerAsnIleLysSerSerL 307

1229 ACATCAACCAAGGCGCGGGCGGTTTGTATTTTGAGGGTAATTTTACGGTC 1278
 307 ysAlaThr......GlyAsnGlyLeuPheSerIleGlySerSerThrIle 321

1279 TCGCCTAAAAACAACGAAACGTGGCAAGGCGCGGGCGTTCATATCAGTGA 1328
 322 LysArgLysIleIleAsnVal......GlyAlaGlyTyrGluAspThrAs 336

1329 TGGCAGTACCGTTACTTGG......AAAGTAAACGGCGTGGCAAACGACC 1372
 336 pAlaValAsnValAlaGlnLeuLysAlaValGluAsnLeuAlaLysArgG 353

1373 GCCTGTCCAAAATCGGC......AAAGGCACGCTGCTGGTTCAAGCCAAA 1416
 353 lnIleThrPheLysGlyAspAspAsnGlyThrGlyValLysLysLysLeu 369

1417 GGGGAAAACCAAGGCTCGGTCAGCGTGGGCGACGGTAAAGTCATCTTAGA 1466
 370 GlyGlu.........ThrLeuThrIleLysGlyGlyGluThrGlnAlaAs 383

1467 TCAGCAGGCGGACGATCAAGGCAAAAAACAAGCCTTTAGTGAAATCGGCT 1516
 383 pLysLeuThrAspAsn.....................AsnAsnIleGlyV 393

1517 TGGTCAGCGGCAGGGGGACG.........GTGCAACTG............ 1545
 393 alValThrAspAsnAsnThrGlyLeuLysValLysLeuAlaLysAsnLeu 409

1546 .................AATGCCGATAATCAGTTCAACCCCGACAAACT 1577
 410 SerGlyLeuGluThrValSerThrLysAsnLeuThrAlaSerGluLysVa 426

1578 CTATTTCGGCTTTCGCGGCGGACGTTTGGATTTGAACGGGCATTCGCTTT 1627
 426 lThrValGlySerGlyAsnAsnThrAlaGluLeuGlnSerGlyGlyLeuT 443

1628 CGTTC..........................CACCGCATTCAAAATACC 1650
 443 hrPheThrProThrThrAsnAlaSerThrAspLysThrValTyrGlyThr 459

1651 GATGAAGGGGCGATGATTGTCAACCACAATCAAGACAAAGAATCCACCGT 1700
 460 AspGlyLeuLysPheThrAspAsnSerAsnThrAlaLeuGluAspThrTh 476

1701 TACCATTACA.................GGCAATAAAGATATTACTACAA 1732
 476 rArgIleThrLysAspLysIleGlyPheSerAsnLysAlaGlyThrValA 493

1733 CCGGCAATAACAACAACTTGGATAGCAAAAAAGAAATTGCCTACAACGGT 1782
 493 spGluAsnLysProTyrLeuAspLysAspLysLeuLysValGlyAsn... 508

1783 TGGTTTGGCGAGAAAGATGCAACCAAAACGAACGGGCGGCTCAATCTGAA 1832
 509 .................SerThrLeuAsnAsnGlyGlyLeuThrValAs 519

1833 TTACCAACCGGAAGAAGCGGATCGCACTTTACTGCTTTCCGGCGGAACAA 1882
 519 n.....................AsnThrIle.........GlyGlySerA 526

1883 AT.....................................TTAAACGGCAAT 1896
 526 snLysGlnIleGlnValGlyAlaAspGlyIleLysPheAlaAspValAsn 542

1897 ATCACGCAAACAAACGGCAAACTGTTTTTCAGCGGCAGACCGACACCGCA 1946
 543 ValAsnValSerAsnAlaAlaLysPheGlyThrThrArgIleThrGluGl 559

1947 CGCCTACAATCATTTAGGAAGCGGGTGGTCAAAAATGGAAGGTATCCCAC 1996
 559 uGluIleGlyPheAlaAspAla..........................A 567

1997 AAGGAGAAATCGTGTGGGACAACGATTGGATCGACCGCACATTTAAAGCG 2046
 567 spGlyLysValAspLysLysSerProTyrLeuAspLysLys......... 580

2047 GAAAACTTCCATATTCAGGGCGGACAAGCGGTGGTTTCCCGCAATGTTGC 2096
 581 .........GlnLeuGlnValGlyGlyValLysIleThrLysAspSerGl 594

2097 CAAAGTGGAAGGCGATTGGCATTTAAGCAATCACGCCCAAGCAGTT.... 2142
 594 yIleAsnAlaGlyAspGlnLysIleSerAsnValLysAspAlaThrAspA 611

2142 ..................................................... 2142
 611 spThrAspAlaValThrTyrLysGlnLeuLysGlnValGlnGlnAspAla 627

2143 .................TTCGGTGTCGCACCGCAT......CAAAGCCA 2168
 628 AspGlyAlaLeuGlnSerPheSerIleArgAspGluLysGlyGlnGluPh 644

2169 CACAATCTGTACACGTTCGGACTGGACGGGTCTGACAAGTTGTACCGAAA 2218
```

```
          ||||||   :::      :::   :::|||  |||      |||   :
 644 eThrIleSerAsnLeuTyrSer...AsnGlyAsnThrProAsnThrPheG 660

2219 AAACCATTACC.........GACGATAAAGTGATTGCTTCATTGAGCAAG 2259
     ::|||||||||         ::::::   :::      |||:::|||:::
 660 luThrIleThrPheAlaGlyGluAsnGlyIle......SerIleSerAsn 674

2260 ACCGACATCAGAGGCAATGTCAGCCTTGCCGATCACGCTCATTTAAATCT 2309
              :::|||:::|||::::::
 675 AspIleAlaLysGlyLysValLysVal....................... 683

2310 CACAGGACTTGCCACACTCAACGGC...........AATCTTAGTGCAG 2347
         |||:::      :::||||||            :::|||::::::|
 684 ....GlyIleAspProIleAsnGlyLeuThrThrProLysLeuThrValG 699

2348 GCGGAGAC...........ACGCACTATACGGTTACGCGCAACGCCACC 2385
     ||:::|||           |||:::   ::::::  :::   |||:::
 699 lySerAspLysAspGlyLysThrGlnLeuValIleGluGlnValAlaSer 715

2386 CAAAACGGCAACCTCAGCCTCGTGGGCAATGCCCAAGCAACATTTAAT.. 2433
        ||||||:::  :::::::::  :::::::::   |||:::
 716 GlyAsnGlyThrLysAsnIleIleArgGlyValSerProThrLeuProSe 732

2434 ......................CAAGCCACATTAAACGGCAACACATCGG 2461
                           ::::::|||  :::|||||||||   :
 732 rIleThrAsnAlaGlyGlyValArgThrThrGluGlnGlyAsnThrIleT 749

2462 CTTCGGAC....................................AAT 2472
     ::||||||                                    |||
 749 hrSerAspGluAspLysSerLysAlaAlaSerIleGlyAspIleLeuAsn 765

2473 GCTTCATTTAATCTAAGCAACAACGCCGTACAAAACGGCAGTCTGACG.. 2520
     ::::::|||||||||:::||||||:::   :::   |||   ::::::
 766 ThrGlyPheAsnLeuLysAsnAsnSerAsnSerValGlyPheValSerTh 782

2521 ..................CTTTCCGACAACGCT...AAGGCAAACGTAA 2548
                       ::::::  ||||||      |||:::|||:
 782 rTyrAsnThrValAspPheIleAspGlyAsnAlaThrThrAlaLysValT 799

2549 GCCATTCCGCACTCAACGGC.................AATGTCTCCCTA 2580
     ::::::::        |||                  :::|||::::::
 799 hrTyrAspGluThrAsnGlnThrSerLysValThrTyrAspValAsnVal 815

2581 GCCGATAAGGCAGTATTCCATTTTGAAAACAGCCGCTTTACCGGAAAAAT 2630
        :::|||::::::             ::::::    |||:::||||||
 816 AspGluLysThrIleGluLeuThrGlyAspAsnGlyLysThrAsnLysIl 832

2631 C...............................AGCGGCGGCAAGGATA 2647
     |                               ::::::||||||   |
 832 eGlyValLysThrThrThrLeuThrThrThrAsnAlaAsnGlyLysAlaT 849

2648 CGGCATTACACTTAAAAGACAGCGAATGGACGCTGCCGTCGGGCACGGAA 2697
     ||    :::         |||::::::      :::   :::
 849 hrAsnPheSerThrThrAspAsnAspAlaLeuValAsnAlaLysAspIle 865

2698 TTAGGCAATTTAAAC................................... 2712
          |||||||||
 866 AlaGluAsnLeuAsnThrLeuAlaLysGluIleHisThrThrLysGlyTh 882

2713 ...........................................CTTG 2716
                                                   |
 882 rAlaAspThrAlaLeuGlnThrPheLysValLysLysAspGlyAlaThrA 899

2717 ACAACGCCACCATTACACTCAATTCCGCCTATCGACACGATGCGGCAGGC 2766
     ||:::   |||||||||:::::::::                     |||
 899 spAspGluThrIleThrValGlyLys..................AspGly 909

2767 GCGCAAACCGGCAGTGCGGCAGATGCGCCGCGCCGCCGTTCGCGCCGTTC 2816
     :::|||:::|||:::::::::
 910 ThrGlnAsnGlyLysThrVal............................ 916

2817 CCTATTATCCGTTACGCCGCCAACTTCGGCAGAATCCCGTTTCAACACGC 2866
                                                |||||||

 917 ..........................................AsnThrL 919

2867 TGACGGTAAACGGCAAATTGAACGGTCAGGGAACATTCCGCTTTATGTCG 2916
     ||   ::::::|||:::   ||||||
 919 euLysLeuLysGlyGlu...AsnGly........................ 926

2917 GAACTCTTCGGCTACCGCAGCGGCAAATTGAAGCTGGCGGAAAGTTCCGA 2966
                                |||   :::|||   ::::::::
 927 ...........................LeuThrValAlaThrAsnLysAs 934

2967 AGGCACTTACACCTTGGCTGTCAACAATACCGGCAACGAACCCGTAAGTC 3016
     :||||||   |||:::::::::|||:::   ::::::
 934 pGlyThrValThrPheGlyIleAsnThrGlnSerGly............ 946

3017 TCGAGCAATTGACGGTAGTGGAAGGAAAAGACAACACACCGCTGTCCGAA 3066
                      :::::::::   |||:::|||   |||::::::
 947 .................LeuLysAlaGlyAspSerThrThrLeuAsnLys 957

3067 .........AATCTTAATTTCACCCTGCAAAACGAACACGTCGATGCCGG 3107
                :::::::::         :::||||||:::::::::::||
 958 AspGlyLeuSerIleLysAsnProAlaSerAsnGluGlnIleGlnValGl 974

3108 CGCA......TGGCGTTATCAGCTTATCCGCAAA................ 3135
     ||||         ::::::      :::   |||
 974 yAlaAspGlyValLysPheAlaLysValAspLysGlyAsnSerSerThrG 991

3136 .....GACGGCGAGTTCCGCCTGCAT........................ 3156
          ||||||   |||:::
 991 lyIleAspGlyThrSerArgIleThrLysAspGlnIleGlyPheThrGly 1007

3157 ...............AATCCGGTCAAAGAACAAGAGCTTTCCGACAAACT 3191
                    :::   :::|||   :::      :::||||||||
1008 AlaAsnGlySerLeuAspThrThrLysProHisLeuThrLysAspLysLe 1024

3192 CGGCAAGGCGGGAGAAACAGAGGCCGCCTTGACGGCAAAACAGGCA.... 3237
     |   |||:::||||||:::|||   :::   |||:::   :::|||
1024 u...LysValGlyGluValGluIleThrAsnThrGlyIleAsnAlaGlyG 1040

3238 .....CAACTTGCCGCCAAACAACAGGCGGAAAAAGACAACGCGCAAAGC 3282
          :::::::::      |||::::::       |||   :::|||:::
1040 lyLysLysIleThrAsnIleGlnSerGly......AspIleThrGlnAsn 1054

3283 CTTGACGCGCTGATTGCGGCCGGGCGC..................AATGC 3314
        :::     :::::::::||||||                    :::
1055 SerAsnAspAlaValThrGlyGlyArgValTyrAspLeuLysThrGluLe 1071

3315 CACCGAAAAGGCAGAAAGTGTTGCCGAACCGGCCCGGCAGGCA 3357
        :::|||   :::|||:::|||:::   |||:::::::::
1071 uGluSerLysIleAsnSerAlaAlaLysThrAlaGlnAsnSer 1085
```

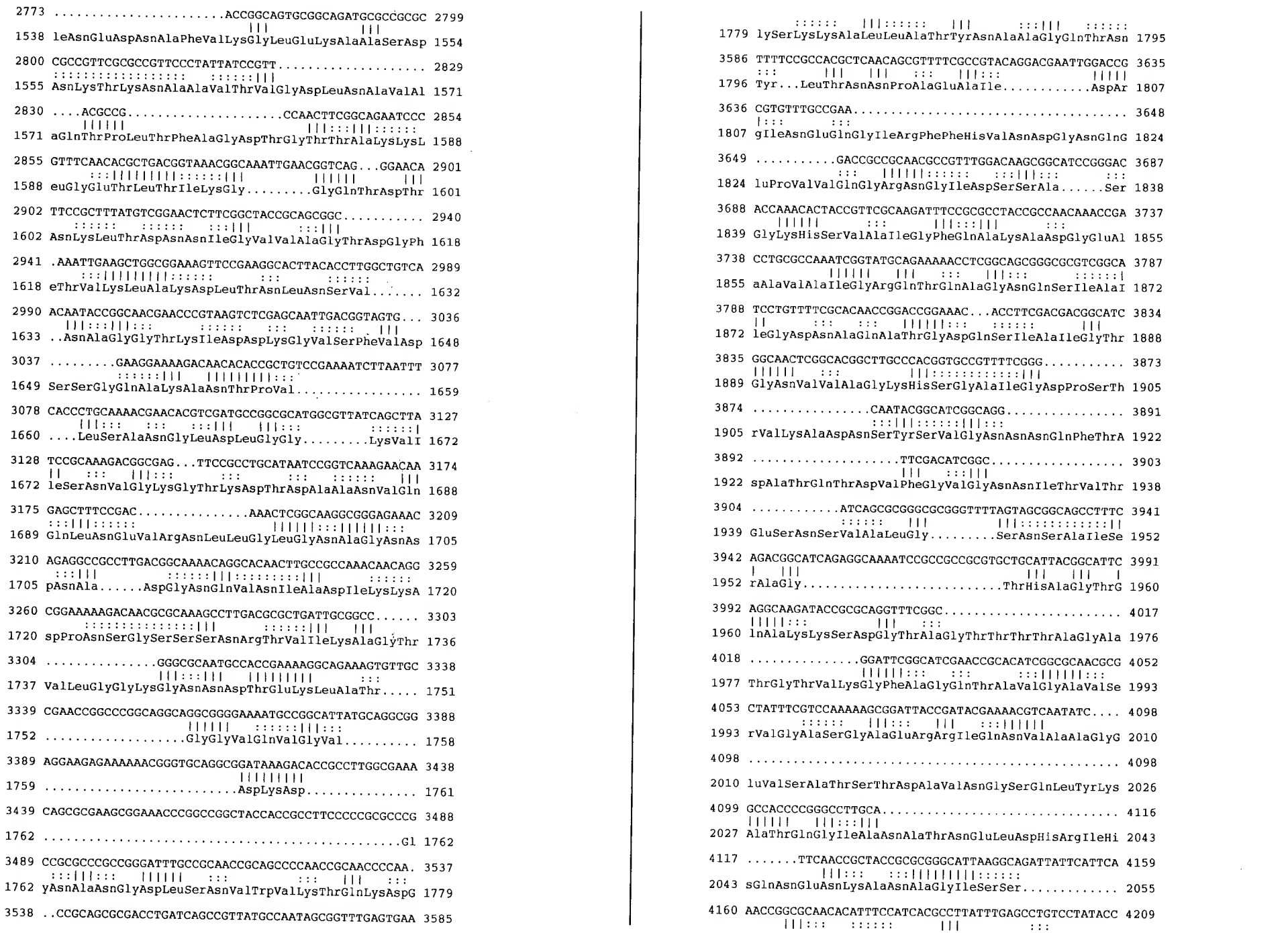
seq_name: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-08-968-685A-10

seq_documentation_block:
; Sequence 10, Application US/08968685A
; Patent No. 6214981
; GENERAL INFORMATION:
; APPLICANT: TUCKER, KENNETH
; APPLICANT: PLOSILA, LAURA
; TITLE OF INVENTION: MORAXELLA CATARRHALIS OUTER MEMBRANE
; TITLE OF INVENTION: PROTEIN-106 POLYPEPTIDE, GENE SEQUENCE AND USES THEREOF
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30

```
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/968,685A
;      FILING DATE:  No. 6214981ember 12, 1997
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Baldwin, Geraldine F.
;      REGISTRATION NUMBER:  31,232
;      REFERENCE/DOCKET NUMBER:  7969-060
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 790-9090
;      TELEFAX:  (212) 869-8864
;      TELEX:  66141 PENNIE
;  INFORMATION FOR SEQ ID NO:  10:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2123 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  peptide
US-08-968-685A-10


alignment_scores:
            Quality:  228.50              Length:   1754
              Ratio:  0.290                Gaps:     88
 Percent Similarity:  44.983   Percent Identity: 19.327

alignment_block:
 US-09-303-518D-653 x US-08-968-685A-10   ..

 Align seg 1/1  to: US-08-968-685A-10  from: 1  to: 2123

     122 GGGCGGGACACACTTATTTCGGCATCAACTACCAATACTATCGCGACTTT 171
         |||  |||  :::      ::::::  :::|||:::
     590 GlyAlaAspGlyIleThrPheThrAspIleSerAsnSer..........  602

     172 GCCGAAAATAAAGGCAAGTTTGCAGTCGGGGCGAAAGATATTGAGG.... 217
             |||      |||:::::::::::::      :::|||  |||
     603 ....LysProGlyAlaGlyIleGluAsnThrThrArgIleThrArgAspG 618

     218 ........TTTACAACAAAAAAGGGGAGTTGGTCGGCAAATCGATGACGA 259
                 |||:::::::::  ||||||  :::|||||||
     618 lyIleGlyPheAlaAsnAsnThrGlySerLeuAspAlaAsn......... 631

     260 AAGCCCCGATGATTGATTTTTCTGTGGTATCGCGTAACGGCGTGGCGGCA 309
         |||||||||  :::      :::   ::::::  :::|||
     632 LysProArg..LeuThrProThrGlyIleAsnAlaGlyGly...LysGlu 646

     310 TTGGCGGGCGATCAATATATTGTGAGCGTGGCACATAACGGCGGCTATAA 359
         |||::::::  |||      ::::::  |||  |||||||||
     647 LeuThrAsnValGlnSerAlaIleAsnProAlaThrAsnGlyGly..... 661

     360 CAATGTTGATTTTGGTGCGGAGGGAAGCAATCCCGATCAGCACCGCTTTT 409
          ::::::||||||      :::   |||:::  :::  ::::::  :
     662 .GlnLeuAspPheMetAsnArgLeuSerThrAlaAsnThrGluLysSerG 678

     410 CTTACCAAATTGTGAAAAGAAATAATTATAAAGCAGGGACTAACGGCCAT 459
         ::         :::  ::::::  |||:::  :::
     678 lySerAlaAlaThrIleLysAspLeuTyrAsnLeuSerGlnValProLeu 694

     460 CCTTATGGCGGCGATTATCATATGCCGCGTTTGCACAAATTTGTCACAGA 509
          ::::::||||||        |||::::::  |||
     695 ThrPheAlaGlyAspThrGly...ProAsnValThrLys..........  706

     510 TGCAGAACCTGTTGAGATGACCAGTTATATGGATGGGTGGAAATACGCTG 559

     706 ..................................................  706

     560 ATTTAAATAAATACCCTGATCGTGTTCGAATCGGAGCAGGCAGACAATAT 609
             |||      :::  :::::::::  :::|||:::
     707 ........LysLeuGlyGluIleLeuLysValLysGlyGlyLysThr... 719

     610 TGGCGGTCTGATGAAGACGAACCCAATAAC.................... 639

           :::|||:::        ||||||
     720 ...ThrAlaAspAspLeuThrLysAsnAsnIleGlyValValAlaAspSe 735

     640 .......CGCGAAAGTTCATATCATATTGCAAGCGCATATTCTTGGCTCG 682
               ::::::  :::      :::|||::::::  |||  |||
     735 rThrAspAsnSerLeuThrValLysLeuAlaLysThrLeuSerAspLeuA 752

     683 TCGGTGGCAATACCTTTGCACAAAATGGATCAGGTGGTGGCACAGTCAAC 732
           :::  ||||||  :::  ::::::|||        ||||||:::
     752 spAlaValAsnThrLysThrLeuThrAlaSerAspLysValThrValAsp 768

     733 TTAGGTAGCGAA......AAAATTAAACATAGCCCATATGGTTTTTTACC 776
           |||::::::      |||:::::::::::::        |||
     769 SerGlyAsnAsnThrAlaLysLeuGlnAsnGlyAspLeuThrPheSerLy 785

     777 AACAGGAGGCTCATTTGGCGACAGTGGCTCACCAATGTTTATCTATGATG 826
           :::  :::        :::  :::        ||||||  :
     785 sGlnAsnThrGlyAlaThrProAlaThrAsnSerLysThrIleTyrGlyV 802

     827 CCCAAAAGCAAAAGTGGTTA...ATTAATGGGGTATTGCAAACAGGCAAC 873
         :::::      |||:::        ||||||:::        |||:::
     802 alAspGlyLeuLysPheThrAspAsnAsnGlyIleAlaLeuAspGlyThr 818

     874 CCCTATATAGGAAAAAGCAAT...GGCTTCCAGCTAGTTCGTAAAGATTG 920
           ||||||  |||::::::  ||||||      :::::::::|||  |
     819 ThrTyrIleThrLysAspLysValGlyPhe......AlaLysGlnAsp.G 833

     921 GTTCTATGATGAAATCT.....................TTGCTGGAG 946
         |||||:::  ||||||                       |||
     833 lySerLeuAspLysSerLysProTyrLeuAspLysAspLysLeuLysVal 849

     947 ATACCCATTCAGTATTCTACGAACCACATCAAAATGGGAAATACTTTTTT 996
               ::::::  :::||||||  |||:::  |||:::      ::
     850 GlyGluValGluIleThrThrAsnGlyIleAsnAlaGlyGlyLysAlaIl 866

     997 AACG............ACAATAATAATGGCGCAGGAAAAATCGATGCCAA 1034
         :|||            |||:::      |||  :::  :::      :
     866 eThrGlyLeuSerAsnThrLeuThrAspAlaThrAsnAlaThrThrGlyH 883

    1035 ACATAAACACTATTCTCTACCTTATAGATTA.......AAAACACGAACC 1077
         :::::  |||  |||  ::::::|||        ||||||||||:::
     883 isVal...ThrGlnLeuGlyIleValAspSerThrAspLysThrArgAla 898

    1078 GTTCAATTGTTTAATGTTTCTTTATCCGAGACAGCAAGAGAACCTGTTTA 1127
         :::::::::  :::|||
     899 AlaSerIleGlyAspVal.............................. 904

    1128 TCATGCTGCAGGTGGGGTCAACAGTTATCGACCCAGACTGAATAATGGAG 1177
                                                |||||  |||
     905 ..................................LeuAsnAlaGlyP 909

    1178 AAAATATTTCCTTTATTGACAAAGGAAAAGGTGAATTGATACTTACCAGC 1227
           |||::::::          :::|||      :::        :::
     909 heAsnLeuLysAsnAsnGlyAspAlaLysAspPheValSerThrTyrAsp 925

    1228 AACATCAACCAAGGCGCGGGCGGTTTGTATTTTGAGGGTAATTTTACGGT 1277
         :::::::::                 :::::::::||||||  |||::
     926 ThrValAsp.................PheIleAsnGlyAsnAlaThrTh 936

    1278 CTCGCCTAAAAACAACGAAACGTGGCAAGGCGCGGGCGTTCATATCAGT. 1326
         ::::       :::   |||:::::::|||  :::      ::::::
     936 rAla.........LysValThrTyrAspGlyLysAlaSerLysValAlaT 950

    1327 ..............GATGGCAGTACCGTTACTTGGAAAGTAAACGGCGTG 1362
                        ||||||:::|||:::        ::::::|||
     950 yrAspValAsnValAspGlyThrThrIle......HisLeuThrGly... 963

    1363 GCAAACGACCGCCTGTCCAAAATCGGC...AAAGGCACGCTGCTGGTTCA 1409
         |||:::  :::  ::::::||||||  |||  |||  |||::::::
     964 AlaAspGlyAsnLysAsnGlnIleGlyValLysThrThrThrLeuThrLy 980

    1410 A......GCCAAAGGGGAAAAC......CAAGGCTCGGTCAGCGTGGGCG 1447
         :      |||||||||:::::      :::  ||||||:::  ||||
```


980 sThrAspAlaLysGlyAspLysAlaIleAsnPheSerValAsnSerGlyA 997
1448 ACGGTAAAGTCATCTTAGATCAGCAGGCGGACGATCAAGGCAAAAAACAA 1497
997 spAspLysAlaLeuIleAsn.....AlaLysAsp 1006
1498 GCCTTTAGTGAAATCGGCTTGGTCAGCGGC.....AGGGG 1532
1007 IleAlaAspAsnLeuAsnThrLeuAlaGlyGluIleArgAsnThrLysGl 1023
1533 GACG.....GTGCAACTGA 1546
1023 yThrAlaAspThrAlaLeuGlnThrPheGlnValLysLysValLysGluA 1040
1547 ATGCCGATAATCAGTTCAACCCCGACAAACTCTATTTCGGCTTTCGCGGC 1596
1040 snGlyAspAspAspAsnAspAlaAspThrIleThrValGlyLysAspAla 1056
1597 GGACGTTTGGATTTGAACGGGCATTCGCTTTCGTTCCACCGCATTCAAAA 1646
1057 LysThrAsnGlnValAsnThrLeuLysLeu..... 1066
1647 TACCGATGAAGGGGCGATGATTGTCAACCACAATCAAGACAAAGAATCCA 1696
1067 .....LysGlyLysAsnGlyLeuAspIleGlnThrAsnLysAspGlyT 1081
1697 CCGTTACCATT.....ACAGGCAATAAA 1719
1081 hrValThrPheGlyIleAsnThrGlnSerGlyLeuLysAlaGlyAsn... 1096
1720 GATATTACTACAACCGGCAATAACAACAACTTGGATAGCAAAAAAGAAAT 1769
1097 .....AsnThrThrLeuAsnAsnAsnGlyLeu...SerIleLysAsnTh 1110
1770 TGCCTACAAC.....GGTTGGTTTGGCGAGAAAGATGCAA 1804
1110 rAlaGlyAsnGluGlnIleGlnValGlyAlaAspGlyValLysPheAlaL 1127
1805 CCAAAACGAACGGGCGGCTCAATCTGAATTACCAACCGGAAGAAGCGGAT 1854
1127 ysValAsnAsnGlyValVal..... 1133
1855 CGCACTTTACTGCTTTCCGGCGGAACAAATTTAAACGGCAATATCACGCA 1904
1134 .....GlyAlaGlyIleAspGlyThrThrArgIl 1143
1905 AACAAACGGCAAACTGTTTTTCAGCGGC.....AGAC 1936
1143 eThrArgAspGluIleGlyPheAlaGlyThrAsnGlySerLeuAspLysS 1160
1937 CGACACCGCAC...GCCTACAATCATTTAGGAAGCGGGTGGTCAAAAATG 1983
1160 erLysProHisLeuSerLysAspGlyIleAsnAlaGlyGlyLysLysIle 1176
1984 GAAGGTATCCCACAAGGAGAAATCGTGTGGGACAACGATTGGATCGACCG 2033
1177 ThrAsnIleGlnSerGlyGluIleAlaGlnAsnSer..... 1188
2034 CACATTTAAAGCGGAAAACTTCCATATTCAGGGCGGACAAGCGGTGGTTT 2083
1189 .....AsnAspAlaValThrGlyGlyLysIleTyr.... 1198
2084 CCCGCAATGTTGCCAAAGTGGAAGGCGATTGGCATTTAAGCAATCACGCC 2133
1199 .....AspLeuLysThrGluLeuGluAsnLysIleSerSerThrAla 1212
2134 CAAGCAGTT.....TTCGGTGTCGCACCGCATCAAAG 2165
1213 LysThrAlaGlnAsnSerLeuHisGluPheSerValAlaAspGluGlnGl 1229
2166 CCACACAATCTGTACACGTTCGGACTGGACGGGTCTG...ACAAGTTGTA 2212
1229 yAsnAsnPheThrValSerAsnProTyrSerSerTyrAspThrSerLysT 1246


2213 CCGAAAAAACCATTACC.....GACGATAAAGTGATTGCTTCATTG 2253
1246 hrSerAspValIleThrPheAlaGlyGluAsnGlyIleThrThrLysVal 1262
2254 AGCAAGACCGACATCAGA..... 2271
1263 AsnLysGlyValValArgValGlyIleAspGlnThrLysGlyLeuThrTh 1279
2272 .....GGCAATGTCAGCCTTG 2287
1279 rProLysLeuThrValGlyAsnAsnAsnGlyLysGlyIleValIleAspS 1296
2288 CCGATCACGCTCATTTAAATCTCACAGGACTTGCCACACTCAACGGCAAT 2337
1296 erGlnAsnGlyGlnAsnThrIleThrGlyLeuSerAsnThrLeuAlaAsn 1312
2338 CTTAGTGCAGGCGGAGACACGCACTATACGGTTACGCGCAACGCCACCCA 2387
1313 Val.....ThrAsnAspLysGlySerValArgThrThrGl 1324
2388 AAACGGCAACCTC.....AGCCTCG 2407
1324 uGlnGlyLysIleIleLysAspGluAspLysThrArgAlaAlaSerIleV 1341
2408 TGGGCAATGCCCAAGCAACATTTAATCAAGCCACATTAAACGGCAACACA 2457
1341 alAspValLeuSerAlaGlyPheAsn.....LeuGlnGlyAsnGly 1354
2458 TCGGCTTCGGACAATGCTTCA.....TTTAATCTAAGCAACAA 2495
1355 GluAlaValAspPheValSerThrTyrAspThrValAsnPheAlaAspGl 1371
2496 CGCCGTACAAAACGGCAGTCTGACGCTTTCCGACAACGCTAAGGCAAAC. 2544
1371 yAsnAlaThrThrAlaLysValThrTyrAspAspThrSerLysThrSerL 1388
2544 2544
1388 ysValValTyrAspValAsnValAspAspThrThrIleGluValLysAsp 1404
2544 2544
1405 LysLysLeuGlyValLysThrThrThrLeuThrSerThrGlyThrGlyAl 1421
2544 2544
1421 aAsnLysPheAlaLeuSerAsnGlnAlaThrGlyAspAlaLeuValLysA 1438
2545 .....GTAAGCCAT...TCCGCACTCAACGGCAATGTCTCCCTA 2580
1438 laSerAspIleValAlaHisLeuAsnThrLeuSerGlyAspIleGlnThr 1454
2581 GCC..... 2583
1455 AlaLysGlyAlaSerGlnAlaAsnSerSerAlaGlyTyrValAspAlaAs 1471
2584 ....GATAAGGCAGTATTC...CATTTTGAAAACAGCCGCTTTACCGGAA 2626
1471 pGlyAsnLysValIleTyrAspSerThrAspAsnLysTyrTyrGlnAlaL 1488
2627 AAATCAGCGGCGGCAAGGATACGGCATTACACTTAAAAGACAGCGAATGG 2676
1488 ysAsnAspGlyThrValAspLysThrLysGluValAlaLysAspLysLeu 1504
2677 ACGCTGCCGTCGGGCACGGAATTAGGCAATTTAAACCTTGACAACGCCAC 2726
1505 ValAlaGlnAlaGlnThrProAspGlyThrLeuAlaGlnMetAsnValLy 1521
2727 CATTACACTCAATTCCGCCTATCGACACGATGCGGCAGGCGCGCAA.... 2772
1521 sSerValIleAsnLysGluGlnValAsnAspAlaAsnLysLysGlnGlyI 1538



```
    2056 .....AlaMetAlaMetAlaSerMetProGlnAlaTyrIleProGlyArg 2070

    4210 GATGCCGCTTCCGGCAAAGTCCGAACGCGCGTCAATACCGCCGTATTGGC 4259
         :::   :::::::|||   :::   |||:::   :::   :::::::::||
    2071 SerMetValThrGlyGlyIleAlaThrHisAsnGlyGlnGlyAlaValAl 2087

    4260 GCAGGATTTCGGCAAA...ACCCGCAGTGCGGAATGG............G 4294
         |      :::::::|||   :::   :::::::::|||          :
    2087 aValGlyLeuSerLysLeuSerAspAsnGlyGlnTrpValPheLysIleA 2104

    4295 GCGTAAACGCCGAAATCAAAGGTTTCACGCTGTCCCTCCACGCTGCCGCC 4344
         ::   :::|||:::   :::|||              |||::::::|||
    2104 snGlySerAlaAspThrGlnGly..............HisValGlyAla 2115

    4345 GCCAAGGGG 4353
         |||   |||
    2116 AlaValGly 2118

seq_name: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-09-268-347-48

seq_documentation_block:
; Sequence 48, Application US/09268347
; Patent No. 6335182
; GENERAL INFORMATION:
;  APPLICANT: Loosmore, Sheena M.
;  TITLE OF INVENTION: RECOMBINANT HAEMOPHILUS INFLUENZAE ADHESIN PROTEINS
;  FILE REFERENCE: 1038-860
;  CURRENT APPLICATION NUMBER: US/09/268,347
;  CURRENT FILING DATE: 1999-03-16
;  NUMBER OF SEQ ID NOS: 54
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 48
;   LENGTH: 2048
;   TYPE: PRT
;   ORGANISM: Haemophilus influenzae
US-09-268-347-48


alignment_scores:
           Quality:  227.00          Length:   1648
             Ratio:  0.311             Gaps:     74
 Percent Similarity:  44.357   Percent Identity: 19.235

alignment_block:
 US-09-303-518D-653 x US-09-268-347-48   ..

 Align seg 1/1  to: US-09-268-347-48  from: 1  to: 2048

     163 CGCGACTTTGCCGAAAATAAAGGCAAGTTTGCAGTC............GG 200
         :::|||   ::::::|||   |||      ::::::           ||
     657 LysAspThrThrLysAsnAlaGlyAlaValSerIleLeuLysLeuLysGl 673

     201 GGCGAAAGATATTGAGGTTTACAACAAAAAAGGGGAGTTGGTC......G 244
         |   :::   :::   |||   :::|||||||        |||      |
     673 yLysAsnGlyLeuThrValAlaThrLysLysAspGlyThrValThrPheG 690

     245 GCAAATCGATGACGAAAGCCCCGATGATTGATTTTTCTGTGGTATCGCGT 294
         ||   |||:::   ::::::      |||      |||:::::::::::::
     690 lyLeuSerGlnAspSerGlyLeuThrIleGlyLysSerThrLeuAsnAsn 706

     295 AACGGCGTGGCGGCATTGGCGGGCGATCAATATATTGTGAGCGTGGCACA 344
         :::|||:::::::::         ::::::   |||         |||::
     707 AspGlyLeuThrValLysAspThrAsnGluGlnIleGlnValGlyAlaAs 723

     345 TAACGGCGGCTATAACAATGTTGATTTTGGTGCGGAGGGAAGCAATCCCG 394
         ::::      ::::::||||||:::           ||||||||||||
     723 nGlyIleLysPheThrAsnValAsn............GlySerAsnProG 736

     395 ATCAGCACCGCTTTTCTTACCAAATTGTGAAAAGAAATAATTATAAAGCA 444
               :::      :::   |||:::      |||
     736 lyThrGlyIleAlaAsnThrAlaArgIleThrArgAsp......LysIle 750

     445 GGGACTAACGGCCATCCTTATGGCGGCGATTATCATATGCCGCGTTTGCA 494

     751 GlyPheAlaGlySerAspGlyAlaValAspThrAsnLysPro........ 764

     495 CAAATTTGTCACAGATGCAGAACCTGTTGAGATGACCAGTTATATGGATG 544
                                                 |||:::|||
     765 .....TyrLeuAsp. 767

     545 GGTGGAAATACGCTGATTTAAATAAATACCCTGATCGTGTTCGAATCGGA 594
                                        |||:::::::::::::
     768 ..........GlnAspLysLeuGlnVal... 773

     595 GCAGGCAGACAATATTGGCGGTCTGATGAAGACGAACCCAATAACCGCGA 644

     773 ..... 773

     645 AAGTTCATATCATATTGCAAGCGCATATTCTTGGCTCGTCGGTGGCAATA 694
                                                  ||||||:
     774 .....GlyAsnV 776

     695 CCTTTGCACAAAATGGATCAGGTGGTGGCACAGTCAACTTAGGTAGCGAA 744
         ::         |||   :::            :::|||   |||::::::
     776 alLysIleThrAsnThrGly............IleAsnAlaGlyGlyLys 788

     745 AAAATTAAACATAGCCCATATGGTTTTTTACCAACAGGAGGCTCATTTGG 794
            |||              |||:::   ||||||      |||:::::
     789 AlaIleThr............GlyLeuSerProThrLeuProSerIleAl 801

     795 CGACAGTGGCTCACCAATGTTT..............ATCTATGATGCCC 829
         :|||::::::|||      :::              |||   |||   :
     801 aAspGlnSerSerArgAsnIleGluLeuGlyAsnThrIleGlnAspLysA 818

     830 AAAAGCAAAAGTGGTTA...ATTAATGGGGTATTGCAAACAGGCAACCCC 876
         ::|||::::::         ||||||   :::|||:::||||||
     818 spLysSerAsnAlaAlaSerIleAsnAspIleLeuAsnThrGlyPheAsn 834

     877 TATATAGGAAAAAGCAATGGCTTCCAGCTAGTTCGTAAAGATTGGTTCTA 926
               ::::::::::|||   ::::::::::|||              ||
     835 LeuLysAsnAsnAsnAsnProIleAspPheValSerThr.........Ty 848

     927 TGATGAAATC...TTTGCTGGAGATACCCATTCAGTATTCTACGAACCAC 973
         ||||   :::   ||||||:::   :::   ::::::            |
     848 rAspIleValAspPheAlaAsnGlyAsnAlaThrThrAlaThrValThrH 865

     974 ATCAAAATGGGAAATACTTTTTTAACGACAATAATAATGGCGCAGGAAAA 1023
         ||                        |||:::   |||   :::::::|||
     865 is.......................AspThrAlaAsnLysThrSerLys 873

    1024 ATCGATGCCAAACATAAACACTATTCTCTACCTTATAGATTAAAAACACG 1073
         :::
     874 Val..... 874

    1074 AACCGTTCAATTGTTTAATGTTTCTTTATCCGAGACAGCAAGAGAACCTG 1123
                  ::::::::::|||:::::::::::|||          :
     875 .........ValTyrAspValAsnValAspAspThr...........T 884

    1124 TTTATCATGCTGCAGGTGGGGTCAACAGTTATCGACCCAGACTGAATAAT 1173
         ::   |||   :::|||
     884 hrIleHisLeuThrGly............................... 889

    1174 GGAGAAAATATTTCCTTTATTGACAAAGGAAAAGGTGAATTGATACTTAC 1223
                                    |||   :::|||          :::   ||
     890 ..................ThrAspAspAsnLysLysLeuGlyValLysTh 900

    1224 CAGCAACATCAACCAAGGCGCGGGCGGTTTGTATTTTGAGGGTAATTTTA 1273
         |::::::::::|||:::   :::::::::                 ||||||:
     900 rThrLysLeuAsnLysThrSerAlaAsnGlyAsnThrAlaThrAsnPheA 917

    1274 CGGTCTCGCCTAAAAACAACGAAACGTGGCAAGGCGCGGGCGTTCATATC 1323
         ::|||:::   ::::::::::::::::      :::|||
     917 snValAsnSerSerAspGluAspAlaLeuValAsnAlaLysAspIleAla 933

    1324 AGTGATGGCAGTACCGTTACTTGGAAAGTAAAC.........GGCGTGGC 1364
         ::::::   :::|||::::::   :::::::::         |||:::||
```

```
 934 GluAsnLeuAsnThrLeuAlaLysGluIleHisThrThrLysGlyThrAl 950

1365 AAACGACCGCCTGTCCAAAATCGGCAAAGGCACGCTGCTGGTTCAAGCCA 1414
     |:::     |||:::                 |||:::   |||:::
 950 aAspThrAlaLeuGln.............ThrPheThrValLysLysV 962

1415 AAGGGGAAAACCAAGGCTCGGTCAGCGTGGGCGACGGTAAAGTCATCTTA 1464
       ||||||:::::::::
 962 alAspGluAsnAsnAsnAlaAsp..................... 969

1465 GATCAGCAGGCGGACGATCAAGGCAAAAAACAAGCCTTTAGTGAAATCGG 1514
     |||   :::|||         |||:::|||:::|||   ::::::::::
 970 AspAlaAsnAlaIleThrValGlyGlnLysAsnAlaAsnAsnGlnValAs 986

1515 CTTGGTCAGCGGCAGGGGGACGGTGCAACTGAATGCCGATAATCAGTTCA 1564
     :   ::::::   :::|||         ||||||      ::::::   |
 986 nThrLeuThrLeuLysGlyGluAsnGlyLeuAsnIleLysThrAspLysA 1003

1565 ACCCCGACAAACTCTAT.............TTCGGCTTTCGCGGCGGA 1599
     ||            :::                |||::::::::::|||
1003 snGlyThrValThrPheGlyIleAsnThrThrSerGlyLeuLysAlaGly 1019

1600 CGTTTGGATTTGAACGGGCATTCGCTTTCGTTC.............. 1632
     :::     ||||||     :::|||||||:::          '
1020 LysSerThrLeuAsnAspGlyGlyLeuSerIleLysAsnProThrGlySe 1036

1633 .CACCGCATTCAAAATACCGATGAAGGGGCGATGATTGTCAACCACAATC 1681
      ::::::||||||        :::|||:::   :::::::::   |||:
1036 rGluGlnIleGlnValGlyAlaAspGlyValLysPheAlaLysValAsnA 1053

1682 AAGACAAAGAATCCACCGTTACCATTACAGGCAATAAAGATATTACTACA 1731
     :::::            :::   |||   |||:::      ||||||
1053 snAsnGlyValValGlyAlaGlyIleAspGlyThrThrArgIleThrArg 1069

1732 ..............ACCGGCAATAACAACAACTTGGATAGCAAAAAAGA 1766
                    ||||||:::|||:::::::||||||::::::|||
1070 AspGluIleGlyPheThrGlyThrAsnGlySerLeuAspLysSerLysPr 1086

1767 AATTGCCTACAACGGTTGGTTTGGCGAGAAAGATGCAACCAAAACGAACG 1816
                              :::|||||||:::   :::::::::|
1086 oHisLeu.................SerLysAspGlyIleAsnAlaGlyG 1097

1817 GGCGGCTCAATCTGAATTACCAACCGGAAGAAGCGGATCGCACTTTACTG 1866
     ||:::         |||   |||
1097 lyLysLysIleThrAsnIleGln........................... 1104

1867 CTTTCCGGCGGAACAAATTTAAACGGCAATATCACGCAAACAAACGGCAA 1916
       ||||||      :::   |||::::::   :::   |||:::||||
1105 ...SerGlyGluIleGlnAlaAsnSerHisAspAlaValThrGlyGlyLy 1120

1917 ACTGTTT.....................................T 1924
     |:::::
1120 sIleTyrAspLeuLysThrGluLeuGluAsnLysIleSerSerThrAlaL 1137

1925 TCAGCGGCAGACCGACACCGCACGCCTACAATCATTTAGGAAGCGGGTGG 1974
       :::::::::   :::   |||   :::::: :::   :::      ||
1137 ysThrAlaGlnAsnSerLeuHisGluPheSer.ValAlaAspGluGlnGl 1153

1975 TCAAAAATGGAAGGTATCCCACAAGGAGAAATCGTGTGGGACAACGATTG 2024
     |:::::::::   ||||||:::       :::|||      |||:::
1153 yAsnAsnPheThrValSerAsnProTyrSerSerTyrAspThrSerLysT 1170

2025 GATCGACCGC....ACATTTAAAGCGGAAAACTTCCATATTCAGGGCGGA 2070
       |||         ||||||   :::||||||             |||
1170 hrSerAspValIleThrPheAlaGlyGluAsn...........GlyIle 1182

2071 CAAGCGGTGGTTTCCCGCAATGTTGCCAAAGTGGAAGGCGATTGGCATTT 2120
       :::   |||:::::::::|||:::::|||      |||
1183 ThrThrLysValAsnLysGlyValValArgValGlyIleAspGlnThrLy 1199

2121 A.....................AGCAATCACGCCCAAGCAG 2140
                            :::|||:::::::::::::
1199 sGlyLeuThrThrProLysLeuThrValGlyAsnAsnAsnGlyLysGlyI 1216

2141 TTTTCGGTGTCGCACCGCATCAAAGCCACACAATCTGTACACGTTCGGAC 2190
     ::          :::   :::   ::::::||||||
1216 leValIleAspSerGlnAsnGlyGlnAsnThrIle.............. 1227

2191 TGGACGGGTCTGACAAGTTGTACCGAAAAAACCATTACCGACGATAAAGT 2240
       |||||||||::::::   |||      ::::::|||:::||||||
1228 ...ThrGlyLeuSerAsn...ThrLeuAlaAsnValThrAsnAspLys.. 1241

2241 GATTGCTTCATTGAGCAAGACCGACATCAGAGGCAATGTC.......... 2280
       :::|||:::      |||:::   :::|||||||:::
1242 ....GlySerValArgThrThrGlu...GlnGlyAsnIleIleLysAspG 1256

2281 ...................AGCCTTGCCGAT......CACGCTCATTTA 2304
                          |||:::::|||         |||   :::
1256 luAspLysThrArgAlaAlaSerIleValAspValLeuSerAlaGlyPhe 1272

2305 AATCTCACAGGACTTGCCACACTCAACGGC..................... 2334
     ||||||   |||           ||||||
1273 AsnLeuGlnGly............AsnGlyGluAlaValAspPheValSe 1285

2335 ..............AATCTTAGTGCAGGCGGAGACACGCACTATACGG 2368
                    |||::::::   |||:::   |||        |
1285 rThrTyrAspThrValAsnPheAlaAspGlyAsnAlaThrThrAlaLysV 1302

2369 TTACGCGCAACGCCACC............................... 2385
     |||||   :::   |||
1302 alThrTyrAspAspThrSerLysThrSerLysValValTyrAspValAsn 1318

2386 .....................................CAAAACGG 2393
                                          ::::::
1319 AspAspThrThrIleGluValLysAspLysLysLeuGlyValLysThrTh 1335

2394 CAACCTCAGCCTCGTGGGC.........AATGCCCAAGCAACATTTAATC 2434
      :::|||:::   :::|||         |||      |||      ||||
1335 rThrLeuThrSerThrGlyThrGlyAlaAsnLysPheAlaLeuSerAsnG 1352

2435 AAGCCACATTAAACGGCAACACATCGGCTTCGGACAATGCTTCATTTAAT 2484
     |||||||   ::::::   ::::::|||||||||   ::::::   :::
1352 lnAlaThrGlyAspAlaLeuValLysAlaSerAspIleValAlaHisSer 1368

2485 CTAAGCAACAACGCCGTACAAAACGGCAGTCTGACGCTTTCCGACAACGC 2534
     |||:::                          |||||||||   :::
1369 LeuAsn.........................ThrLeuSerGlyAspIl 1376

2535 TAAGGCAAACGTAAGCCATTCCGCACTCAACGGCAATGTCTCCCTAGCCG 2584
      ::::::      :::   |||      |||:::::::::::   :::|
1376 eGlnThrAlaLysGlyAlaSerGlnAlaAsnAsnSerAlaGlyTyrValA 1393

2585 AT............AAGGCAGTATTCCAT......TTTGAAAACAGCCGC 2616
     ||            |||   |||::::::          :::|||:::
1393 spAlaAspGlyAsnLysIleValIleTyrAspSerThrAspAsnLysTyr 1409

2617 TTTACCGGAAAAATCAGCGGCGGCAAGGATACGGCATTACACTTAAAAGA 2666
     :::   :::|||   :::|||       |||   :::   ::::::
1410 TyrGlnAlaLysAsnAspGlyThrValAspLysThrLysGluValAlaLy 1426

2667 CAGCGAATGGACGCTGCCGTCGGGCACGGAATTAGGCAATTTAAACCTTG 2716
      ::::::   :::       :::   |||      |||:::|||
1426 sAspLysLeuValAlaGlnAlaGlnThrProAspGlyThrLeuAlaGlnM 1443

2717 ACAACGCCACCATTACACTCAATTCCGCCTATCGACACGATGCGGCAGGC 2766
       |||:::      ::::::|||:::          :::||||||
1443 etAsnValLysSerValIleAsnLysGluGlnValAsnAspAlaAsnLys 1459

2767 GCGCAA.........................ACCGGCAGTGCGGCAGA 2789
       |||                           |||
1460 LysGlnGlyIleAsnGluAspAsnAlaPheValLysGlyLeuGluLysAl 1476

2790 TGCGCCGCGCCGCCGTTCGCGCCGTTCCCTATTATCCGTT.......... 2829
       |||      :::::::::::::::::   :::::::|||
1476 aAlaSerAspAsnLysThrLysAsnAlaAlaValThrValGlyAspLeuA 1493
```

```
2830 .............ACGCCG.....................CCAACTTCG 2844
                  ||||||                       |||:::
1493 snAlaValAlaGlnThrProLeuThrPheAlaGlyAspThrGlyThrThr 1509

2845 GCAGAATCCCGTTTCAACACGCTGACGGTAAACGGCAAATTGAACGGTCA 2894
     |||::::::      :::||||||||||:::::|||          |||||
1510 AlaLysLysLeuGlyGluThrLeuThrIleLysGly.........GlyGl 1523

2895 G...GGAACATTCCGCTTTATGTCGGAACTCTTCGGCTACCGCAGCGGC. 2940
     |      |||   ::::::   ::::::   :::|||      :::|||
1523 nThrAspThrAsnLysLeuThrAspAsnAsnIleGlyValValAlaGlyT 1540

2941 ...........AAATTGAAGCTGGCGGAAAGTTCCGAAGGCACTTACACC 2979
                 :::||||||||||:::::      :::      :::
1540 hrAspGlyPheThrValLysLeuAlaLysAspLeuThrAsnLeuAsnSer 1556

2980 TTGGCTGTCAACAATACCGGCAACGAACCCGTAAGTCTCGAGCAATTGAC 3029
     :::        |||:::|||:::      ::::::   :::   :::::
1557 Val.........AsnAlaGlyGlyThrLysIleAspAspLysGlyValSe 1570

3030 GGTAGTG............GAAGGAAAAGACAACACACCGCTGTCCGAAA 3067
     :   |||             ::::::|||   ||||||||||:::
1570 rPheValAspSerSerGlyGlnAlaLysAlaAsnThrProVal....... 1584

3068 ATCTTAATTTCACCCTGCAAAACGAACACGTCGATGCCGGCGCATGGCGT 3117
                |||:::   :::   :::|||   |||:::
1585 ..............LeuSerAlaAsnGlyLeuAspLeuGlyGly..... 1594

3118 TATCAGCTTATCCGCAAAGACGGCGAG...TTCCGCCTGCATAATCCGGT 3164
        ::::::|||   :::   |||:::      :::      :::   ::
1595 ...LysValIleSerAsnValGlyLysGlyThrLysAspThrAspAlaAl 1610

3165 CAAAGAACAAGAGCTTTCCGAC..............AAACTCGGCAAGG 3199
     ::::   |||:::|||::::::                  ||||||:::|
1610 aAsnValGlnGlnLeuAsnGluValArgAsnLeuLeuGlyLeuGlyAsnA 1627

3200 CGGGAGAAACAGAGGCCGCCTTGACGGCAAAACAGGCACAACTTGCCGCC 3249
     ||||||::   :::|||          ::::::|||::::::::|||
1627 laGlyAsnAspAsnAla......AspGlyAsnGlnValAsnIleAlaAsp 1641

3250 AAACAACAGGCGGAAAAAGACAACGCGCAAAGCCTTGACGCGCTGATTGC 3299
        ::::::      ::::::::::::::::|||      ::::::|||
1642 IleLysLysAspProAsnSerGlySerSerSerAsnArgThrValIleLy 1658

3300 GGCC.....................GGGCGCAATGCCACCGAAAAGGCAG 3328
      |||                     |||:::|||   |||||||||
1658 sAlaGlyThrValLeuGlyGlyLysGlyAsnAsnAspThrGluLysLeuA 1675

3329 AAAGTGTTGCCGAACCGGCCCGGCAGGCAGGCGGGGAAAATGCCGGCATT 3378
        :::                          ||||||   ::;:::|||:::
1675 laThr.......................GlyGlyIleGlnValGlyVal 1683

3379 ATGCAGGCGGAGGAAGAGAAAAAACGGGTGCAGGCGGATAAAGACACCGC 3428
                                             |||||||||
1684 ....................................AspLysAsp..... 1686

3429 CTTGGCGAAACAGCGCGAAGCGGAAACCCGGCCGGCTACCACCGCCTTCC 3478

1686 ..................................................... 1686

3479 CCCGCGCCCGCCGCGCCCGCCGGGATTTGCCGCAACCGCAGCCCCAACCG 3528
                ::::|||:::   ||||||   :::      :::
1687 ........GlyAsnAlaAsnGlyAspLeuSerAsnValTrpValLysThr 1700

3529 CAACCCCAA...CCGCAGCGCGACCTGATCAGCCGTTATGCCAATAGCGG 3575
     |||   :::      ::::::   |||::::::   |||      :::||
1701 GlnLysAspGlySerLysLysAlaLeuLeuAlaThrTyrAsnAlaAlaGl 1717

3576 TTTGAGTGAATTTTCCGCCACGCTCAACAGCGTTTTCGCCGTACAGGACG 3625
     |   :::::::::      |||   |||   :::   |||:::
1717 yGlnThrAsnTyr...LeuThrAsnAsnProAlaGluAlaIle....... 1730

3626 AATTGGACCGCGTGTTTGCCGAA.......................... 3648

                |||||||:::      :::
1731 .....AspArgIleAsnGluGlnGlyIleArgPhePheHisValAsnAsp 1745

3649 ...................GACCGCCGCAACGCCGTTTGGACAAGCGG 3677
                           :::   ||||||::::::   :::|||::
1746 GlyAsnGlnGluProValValGlnGlyArgAsnGlyIleAspSerSerAl 1762

3678 CATCCGGGACACCAAACACTACCGTTCGCAAGATTTCCGCGCCTACCGCC 3727
     :      :::   ||||||      :::      |||:::|||      :
1762 a......SerGlyLysHisSerValAlaIleGlyPheGlnAlaLysAlaA 1777

3728 AACAAACCGACCTGCGCCAAATCGGTATGCAGAAAAACCTCGGCAGCGGG 3777
     ::                    ||||||   |||   :::   |||:::
1777 spGlyGluAlaAlaValAlaIleGlyArgGlnThrGlnAlaGlyAsnGln 1793

3778 CGCGTCGGCATCCTGTTTTCGCACAACCGGACCGGA.............. 3813
        :::::::|||      :::   :::   ||||||
1794 SerIleAlaIleGlyAspAsnAlaGlnAlaThrGlyAspGlnSerIleAl 1810

3813 ..................................................... 3813

1810 aIleGlyArgThrAsnValValAlaGlyLysHisSerGlyAlaIleGlyA 1827

3814 ......................AACACCTTCGACGACGGCATCGGCAAC 3840
                            |||::::::      ::::::||||||
1827 spProSerThrValLysAlaAspAsnSerTyr......SerValGlyAsn 1841

3841 TCGGCACGGCTTGCCCACGGTGCCGTTTTCGGGCAATACGGCATCGGCAG 3890
     :::   :::::::::   ::::::              :::|||:::|||::
1842 AsnAsnGlnPheThrAspAlaThrGlnThrAspValPheGlyValGlyAs 1858

3891 GTTCGACATCGGCATCAGCGCGGGCGCGGGTTTTAGTAGCGGCAGCCTTT 3940
     :   :::|||   ::::::   :::   :::   :::   ||||||   |
1858 n...AsnIleThrValThrGluSerAsnSerValAlaLeuGlySerAsnS 1874

3941 CAGACGGCATCAGAGGCAAAATCCGCCGCCGCGTGCTGCATTACGGCATT 3990
     ||              |||                     |||   |||
1874 erAlaIleSerAlaGly.................ThrHisAlaGlyThr 1884

3991 CAGGCAAGATACCGCGCAGGTTTCGGC..................... 4017
     |||||||::         |||   :::
1885 GlnAlaLysLysSerAspGlyThrAlaGlyThrThrThrThrAlaGlyAl 1901

4018 ..............GGATTCGGCATCGAACCGCACATCGGCGCAACGC 4051
                    |||||||:::   :::      :::|||||||:::
1901 aThrGlyThrValLysGlyPheAlaGlyGlnThrAlaValGlyAlaValS 1918

4052 GCTATTTCGTCCAAAAAGCGGATTACCGATACGAAAACGTCAATATC... 4098
            ::::::   |||:::   |||   :::|||||||
1918 erValGlyAlaSerGlyAlaGluArgArgIleGlnAsnValAlaAlaGly 1934

4098 ..................................................... 4098

1935 GluValSerAlaThrSerThrAspAlaValAsnGlySerGlnLeuTyrLy 1951

4099 ......................GCCACCCCGGGCCTTGCATTCAACCGCT 4126
                             ||||||      |||      :::   :
1951 sAlaThrGlnSerIleAlaAsnAlaThrAsnGluLeuAspHisArgIleH 1968

4127 ACCGCGCGGGCATTAAGGCAGATTATTCATTCAAACCGGCGCAACACATT 4176
     :::::         ||||||:::   :::::::::   |||:::   :::
1968 isGlnAsnGluAsnLysAlaAsnAlaGlyIleSerSerAlaMetAlaMet 1984

4177 TCCATCACGCCTTATTTGAGCCTGTCCTATACCGATGCCGCTTCCGGCAA 4226
     :::      |||         :::         :::   ::::::|||
1985 AlaSerMetProGlnAlaTyrIleProGlyArgSerMetValThrGlyGl 2001

4227 AGTCCGAACGCGCGTCAATACCGCCGTATTGGCGCAGGATTTCGGCAAA. 4275
      :::   |||:::   :::   :::::::::|||      :::::::|||
2001 yIleAlaThrHisAsnGlyGlnGlyAlaValAlaValGlyLeuSerLysL 2018

4276 ..ACCCGCAGTGCGGAATGG............GGCGTAAACGCCGAAATC 4311
       :::   :::::::::|||           :::   :::|||:::
```

```
     2018 euSerAspAsnGlyGlnTrpValPheLysIleAsnGlySerAlaAspThr 2034

     4312 AAAGGTTTCACGCTGTCCCTCCACGCTGCCGCCGCCAAGGGG 4353
          :::|||               |||::::::||||||   |||
     2035 GlnGly.............HisValGlyAlaAlaValGly 2043

 seq_name: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-09-413-814-79

 seq_documentation_block:
 ; Sequence 79, Application US/09413814
 ; Patent No. 6225064
 ; GENERAL INFORMATION:
 ;  APPLICANT: Gesellschaft fuer Biotechnologische Forschung mbH
 ;  APPLICANT: Bristol-Myers Squibb, Co.
 ;  APPLICANT: Beyer, Stefan
 ;  APPLICANT: Bloecker, Helmut
 ;  APPLICANT: Brandt, Petra
 ;  APPLICANT: Cino, Paul M
 ;  APPLICANT: Dougherty, Brian A
 ;  APPLICANT: Goldberg, Steven L
 ;  APPLICANT: Hofle, Gerhard
 ;  APPLICANT: Mueller, Joachim
 ;  APPLICANT: Reichenbach, Hans
 ;  TITLE OF INVENTION: DNA sequences for enzymatic synthesis of polyketide or
 ;  TITLE OF INVENTION: heteropolyketide compounds
 ;  FILE REFERENCE: PCT/US 99/23535
 ;  CURRENT APPLICATION NUMBER: US/09/413,814
 ;  CURRENT FILING DATE: 1999-10-07
 ;  EARLIER APPLICATION NUMBER: DE 198 46 493.2
 ;  EARLIER FILING DATE: 1998-10-09
 ;  NUMBER OF SEQ ID NOS: 107
 ;  SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 79
 ;   LENGTH: 1213
 ;   TYPE: PRT
 ;   ORGANISM: Sorangium cellulosum
 US-09-413-814-79


 alignment_scores:
             Quality:  211.00            Length:    693
               Ratio:  0.751               Gaps:     33
  Percent Similarity:  40.548   Percent Identity: 24.820

 alignment_block:
  US-09-303-518D-653 x US-09-413-814-79   ..

  Align seg 1/1  to: US-09-413-814-79  from: 1  to: 1213

     2562 CAACGGCAATGTCTCCCTAGCCGATAAGGCAGTATTCCATTTTGAAAACA 2611
          :::|||::::::   |||         |||:::   |||
        5 ArgArgArgAlaAlaProHisLeuArgGlyAlaArgPro.......... 17

     2612 GCCGCTTTACCGGAAAAATCAGCGGCGGCAAGGATACGGCATTACACTTA 2661
                  |||:::::::::|||||   |||   ::::::
       18 .......AlaArgArgGlyAspArgArgGlyGlyLeuAlaLeuArgSerA 32

     2662 AAAGACAGCGAATGGACGCTGCCGTCGGGCACGGAATTAGGCAATTTAAA 2711
            |||:::|||               |||||||||
       32 spArgGluArgAlaArgArgArg...GlyHisGly.............. 42

     2712 CCTTGACAACGCCACCATTACACTCAATTCCGCCTATCGACACGATGCGG 2761
                                     :::               ||
       43 ............................GluGlyLeuGlyAlaGlyGl 49

     2762 CAGGCGCGCAAACCGGCAGTGCGGCAGATGCGCCGCGCCGCCGTTCGCGC 2811
          ||||||||||:::|||   :::   |||         |||      :::|
       49 yArgArgAlaArgArgAlaAlaArgArgArgArgLeuProAlaProGlyP 66

     2812 CGTTCCCTATTATCCGTTACGCCGCCAACTTCGGCAGAATCCCGTTTCAA 2861
          ||   |||
       66 roSerPro..................................... 68
```

```
     2862 CACGCTGACGGTAAACGGCAAATTGAACGGTCAGGGAACATTCCGCTTTA 2911
                                         |||
       69 ........................ProArgGlyAlaProGlnAlaPr 76

     2912 TGTCGGAACTCTTCGGCTACCGCAGCGGCAAATTGAAGCTGGCGGAAAGT 2961
          |||:::   |||   |||   |||          :::||||||
       76 oProGlyAlaArgArgGlyProProArgProHisAlaValGlyGly.... 91

     2962 TCCGAAGGCACTTACACCTTGGCTGTCAACAATACCGGCAACGAACCCGT 3011
            ||||||      :::|||:::      :::         |||   |||
       92 ..ArgArgAspAspArgLeuAlaArgGlyHisArgAlaProArgArgArg 107

     3012 AAGTCTCGAGCAATTGACGGTAGTGGAAGGAAAAGACAACACACCGCTGT 3061
          ::::::||||||:::      ::::::|||:::
      108 ArgGlyArgAlaValAlaArgAlaAlaArgArgAla.............. 119

     3062 CCGAAAATCTTAATTTCACCCTGCAAAACGAACACGTCGATGCCGGCGCA 3111
                           ||||||   |||   ::::::   |||
      120 ..................ProAlaAlaArgArgLysHisGlyArg.... 128

     3112 TGGCGTTATCAGCTTATCCGCAAAGACGG.............CGAGTTCCG 3149
                          |||      |||           ||||||
      129 .................ProGlyLeuArgAspLeuHisValArgValHis 139

     3150 CCTGCATAATCCGGTCAAAGAACAAGAGCTTTCCGACAAACTCGGCAAGG 3199
            |||   :::|||:::      |||      |||:::   ::::::||
      140 GlyAlaAlaGlnGlyArgAspAspArgProProArgArgGlyGluHisGl 156

     3200 CGGGAGAAACAGAGGCCGCCTTGACGGCAAAACAGGCACAACTTGCCGCC 3249
          |   |||::::::      |||                          |||:
      156 yProArgHisGlnProProLeu.......................ArgA 165

     3250 AAACAACAGGCGGAAAAAGACAACGCGCA.............AAGCCT 3284
          ::      ||||||        ||||||              :::|||
      165 rgArgProGlyGlyProGlyAlaArgAlaLeuValAlaGluLeuArgPro 181

     3285 TGACGC............GCTGATTGCGGCCGGGCGCAATGCCACCGAAA 3322
                                |||:::   |||   :::   ::|||
      182 ValGlyLeuArgArgValArgAspAlaArgArgArgArgArgArgArgAs 198

     3323 AGGCAGAAAGTGTTGCCGAACCGGCCCGGCAGGCAGGCGGGGAAAATGCC 3372
            |||          |||  |||   :::|||      :::::::::|
      198 pProArgProTyrProArgLeuGly...SerGlyAlaLeuAlaArgAlaA 214

     3373 GGCATTATGCAGGCGGAGGAAGAGAAAAAACGGGTGCAGGCGGATAAAGA 3422
          ||         :::|||   |||   :::   |||   |||:::
      214 rgGlyAlaArgAlaGlyAspArgValGluLeuGlyProGlyAlaAspGly 230

     3423 CACCGCCTTGGCGAAACAGCGCGAAGCGGAAACCCGGCCGGCTACCACCG 3472
                |||:::::   |||   |||:::|||   |||
      231 AspAlaHisGlyArgValProArgArgGlyArgPro...GlyAlaValVa 246

     3473 CCTTCCCCCGCGCCCGCCGCGC.........CCGCCGGGATTTGCCGCAA 3513
          |||||||||       :::            |||:::   |||
      246 lAlaProProArgHisAspGluArgArgLeuAspProAlaGluAlaProA 263

     3514 CCGCAGCCCCAACCGCAACCCCAACCGCAGCGCGACCTGATCAGCCGTTA 3563
            :::   |||   :::   |||   |||   |||
      263 rgSerHisProArgGlyLeuProArgAlaProArg.............. 274

     3564 TGCCAATAGCGGTTTGAGTGAATTTTCCGCCACGCTCAACAGCGTTTTCG 3613
                                      |||:::   ::::::|||   ||
      275 ........................ArgGluProArgArgArgAspAr 282

     3614 CCGTACAGGACGAATTGGACCGCGTGTTTGCCGAAGACCGCCGCAACGCC 3663
          |               |||      :::      |||||||||:::|||
      282 g..............GlyValAspLeuValAspArgProProAspArg. 293

     3664 GTTTGGACAAGCGGCATCCGGGACACCAAACACTACCGTTCGCAAGATTT 3713
                  |||:::||||||                    |||:::
      294 ........GlyArgArgProGly..............ValAlaGlnHis 302

     3714 CCGCGCCTACCGCCAACAAACCGACCTGCG..................... 3743
```

```
     |||      ||||||:::::::::||||||
 303 ProLeuArgProProAlaArgGluProAlaHisLeuArgAlaArgArgGl 319

3744 .........................CCAAATCGGTATGCAGAAAAACCTC 3768
                               |||:::|||            |||:
 319 yAlaGlyAlaValSerAspArgGlyProArgArg.........AspProH 333

3769 GGCAGCGGGCGCGTCGGCATCCTGTTTTC..................... 3797
     :::::|||   ||||||||:::     :::
 333 isArgArgAspArgArgArgAlaArgIleLeuAlaArgArgGlyAlaAsp 349

3798 ......GCACAACCGGACCGGAAACACCTTCGACGA.........CGGCA 3832
           |||   |||:::      |||   ||||||         :::
 350 ProGlyAlaValProGluAlaProHisAspArgArgAlaAlaValGlnAs 366

3833 TCGGCAACTCGGCACGGCTTGCCCACGGTGCCGTTTTCGGGCAATACGG. 3881
      |||:::   |||             |||          ||||||   |||
 366 pArgArgProGlyProLeuLeuArgArgArgAspHisArgAlaAlaArgA 383

3882 ..CATCGG.................CAGGTTCGACATCGGCATCAGCGC 3911
       ||||||                 ::::::   ||||||   :::|||
 383 laHisArgProSerGlyGluAspProArgLeuProHisArgAlaArgArg 399

3912 GGGCGCGGGTTTTAGTAGCGGCAGCCTTTCAGACGGCATCAGAGGCAAAA 3961
        ||||||              |||:::   |||         :::::
 400 AspArgGlyArgProArgAlaAlaProLeuGlyArgAlaGlyGlyArgGl 416

3962 TCCGCCGCCGCGTGCTGCATTACGGCATTCAGGCAAGATACCGCGCAGGT 4011
     :         |||:::::::   |||::::::          |||   |||
 416 yGlyGluAspArgSerValGlyArgGluAla.........ProGlyArgV 430

4012 TTCGGCGGATTCGGCATCGAACCGCACATCGGCGCAACGCGCTATTTCGT 4061
       ||||||   |||:::|||          ||||||          :::|||
 430 alArgArg...ArgArgArgArgArgTrpArgArgAlaProArgLeuArg 445

4062 CCAAAA......AGCGGATTACCGATACGAAAACGTCAATATCGCCACCC 4105
     |||:::      :::|||:::        :::          |||   ||
 446 ProGluGluAlaAlaGlyValHisAspProArgGlyGlyArgArgProPr 462

4106 CGGGCCTTGCATTCAACCGCTACCGCGCGGGCATTAAGGCAGATTATTCA 4155
     ||||   :::   :::          :::|||          |||
 462 oGlyAlaAlaAlaGluArgGluArgGlnGlyGlyProArgArg....... 476

4156 TTCAAACCGGCGCAACACATTTCCATCACGCCTTATTTGAGCCTGTCCTA 4205
          :::|||:::   :::   :::   |||
 477 ..ProAlaGlySerArgArgGlyArgAlaAla.................. 486

4206 TACCGATGCCGCTTCCGGCAAAGTCCGAACGCGCGTCAATACCGCCGTAT 4255
        |||         |||   :::   :::   |||       :::|||::
 487 ...ArgArgGlyArgArgProAlaAspArgAspArgAlaAlaHisArgLe 502

4256 TGGCGCAGGATTTCGGCAAAACCCGCAGTGCGGAATGGGGCGTAAACGCC 4305
     :||||||         ::::::      :::|||   ||||||:::|||
 502 uGlyAla.........ArgGlyGlyLeuAlaGlyArgGlyArgArgArgA 516

4306 GAAATCAA............AGGTTTCACGCTGTCCCTCCACGCTGCCGC 4343
       :::|||            |||:::|||   :::   ||||||:::|||
 516 spArgGlnProLeuArgAlaArgLeuHisLeuAlaAlaProArgAlaArg 532

4344 ..............................................CG 4345
                                                    ||
 533 AlaAlaProAlaArgArgAlaHisArgGlyAlaArgAlaArgArgGlyAr 549

4346 CCAAGGGGCC..................................... 4355
     |:::||||||
 549 gArgGlyAlaGlyArgValAlaAspArgProValProValProAspHisA 566

4356 ..GCAATTGGAAGCGCAGCACAGCGCGGG 4382
       |||   |||:::||||||:::||||||
 566 rgAlaThrGlyAlaAlaAlaArgArgGly 575
```